IMPLEMENTABLE MULTI-DIMENSIONAL INVERSE SCATTERING THEORY

A thesis presented for the degree of
Doctor of Philosophy
in Electrical & Electronic Engineering
in the
University of Canterbury,
Christchurch, New Zealand.

by
David Guan Hock Tan
B. Sc. (Hons 1)
February 1988

Abstract

Mathematical and engineering aspects of direct and inverse scattering and diffraction problems posed in more than one dimension are considered. A unified approach is introduced. Descriptions of scalar linear wave motion that are commonly invoked when treating inverse problems are summarised and extended. Several original contributions (developed in association with others) to branches of inverse theory are presented. Two aspects of computed tomography (CT) are treated : first, previously neglected consequences of the inevitable sampling of data are examined; second, account is taken of finite (and variable) detector resolution as well as attenuation (the coefficient being assumed constant, as is common) of the radiation on its passage through the body to the detectors, for single photon emission CT. A generalised volume source formulation of scattering/diffraction is developed; it is shown to lead to a sequence of useful approximate formulations which are potentially suitable bases for inversion algorithms. Two particular implementations of the Newton-Kantorovich approach to inverse scattering (in which the form of the scatterer is iteratively refined, with the direct problem being solved at each iteration) are developed : first; the null-field method (or extended boundary condition) is treated; second, a numerical algorithm is devised and implemented for a global inverse theory (previously formulated only analytically). Four aspects of Fourier phase retrieval are investigated : first, the effect of choice of image support for complex images is examined; second, it is shown in an inverse scattering context, that extra information, likely to be available in practice, permits complex-valued image-forms to be readily recovered; third, an extension of the electron-microscopal Gerchberg-Saxton algorithm, adapted for the radio engineering problem of deducing an antenna's aperture distribution from only the magnitude of its far-field radiation pattern, is presented; fourth, a reconstruction is reported of a structure consistent with one of the widely-heralded recently-recorded (by Shechtman) electron diffraction patterns suggestive of quasi-crystalline specimens not exhibiting crystalline symmetry.

Acknowledgements

I am privileged to have been supervised by Professor R. H. T. Bates whose own achievements and accolades are inspiration for his many students. He has expected nothing less than my best effort, and always given his in return. I also acknowledge the opportunities he has engineered, both during the course of my Ph.D. and for my future. This document would have been better had I paid more heed to his suggestions. Sincerest thanks, also, to my associate supervisor, Dr D. J. N. Wall, Department of Mathematics, for his thorough and patient guidance, particularly when the going got tough.

Postgraduate research is an exacting discipline, and I am fortunate to have observed several of its exponents, namely Drs W.R. Fright, K. L. Garden, A. M.Sinton, and T. S. Yeo.

I am particularly indebted to N. P. Brieseman, T. J. Connolly, B. L. K. Davey, P. H. Gardenier, R. G. Lane, C. A. Lim, R. D. Murch and J. X. Qu who have influenced considerably the material contained in this thesis. I thank the Department of Electrical and Electronic Engineering as a whole, for providing facilities and an atmosphere conducive to intensive study. I am pleased to take this opportunity to thank Dr V. Kerdemelidis, A. N. Vernon, and W. K. Kennedy, for encouraging me to maintain good physical health these last four years, and my long-time friends W. A. Earl, T. F. Robinson, and R. W.Simpson.

The award of a University Grants Committee postgraduate scholarship in 1984, the Charles Cook, Warwick House, Memorial Scolarship for 1984, and an Australian Guaranteed Corporation Young Achievers' Award 1987, are gratefully acknowledged.

The demands on daily routine of my studies have been borne admirably by my parents, sister Audrey, and brother Michael, who have sacrificed selflessly, silently, solidly, and always been supportive, throughout my whole life.

Last, but not least, I thank Emeritus Professor G. M. Petersen, Department of Mathematics, University of Canterbury, whose encouragement began my journey in science four years earlier than it might have otherwise. To him, I humbly dedicate this thesis.

Preface

My goal in writing this thesis has been to produce a systematic overview of inverse scattering theory, both approximate and exact, and to incorporate within it the original contributions that I have made (together with others) during my Ph.D. studies. Because an integral part of, and usually a precursor to, a study of inverse problems is examination of associated direct problems, I have also included detailed discussions of such problems in this thesis.

The contents of each chapter of the thesis are successively summarised in the following eight paragraphs. The original contributions are presented as appendices to Chapters 3, 5, and 6. Each appendix is in the form of either a reprint of an already published journal paper, or a preprint of a paper that has recently been submitted. These papers are listed at the end of this preface and are identified (in this preface) by numbers in square brackets.

Chapter 1 serves to introduce, in general descriptive terms, the types of inverse problems considered in this thesis, and their associated direct problems. Basic terminology is defined, and three major tasks involved in treating an inverse scattering problem are identified. Chapters 2, 3 and 4, and 5 and 6, examine each of these in turn. Inverse problems suffer from "ill-posedness", and the consequences for computational solutions are indicated.

Many physical processes manifest themselves as scalar linear wave motion under a wide, but not completely encompassing, range of scientifically and technologically interesting situtations. So, such wave motion serves as an effective basis for treating the respective direct and inverse problems in a unified manner. Chapter 2 derives the time-domain wave equations resulting from mathematical models of several of the more ubiquitous physical processes. After transforming to the (temporal) frequency-domain, it transpires that these equations can then be manipulated to a common form. The quantities of interest in inverse scattering problems are represented by a generalised constitutive parameter, appearing as a coefficient in a partial differential equation. There appears to be a preference in the literature to pose inverse scattering problems in the frequency-, rather than the time-, domain and some reasons for this are suggested.

Chapter 3 introduces some concepts of wave propagation that prove useful when examining direct and inverse scattering problems. Although many more techniques have been devised for particular direct problems (notably, the geometrical theory of diffraction), ways of incorporating them into inverse problems have yet to be found, and for this reason they are neglected in this thesis. An extension [9] of the volume-source formulation of scattering is included as Appendix 3-A.

Various approaches to direct problems that can be incorporated into formulations of the corresponding inverse problems are examined in Chapter 4. Important ap-

proximate schemes are included. Also emphasised are tractable numerical approaches. Appendix 3-A is pertinent to this chapter.

It is rare that a worthwhile technological or scientific problem succumbs to straightforward unqualified solution. Often, the available data are sufficiently limited or distorted versions of idealised data that preprocessing techniques are required to massage the data into a form suitable for an inversion algorithm. Several such problems are discussed in Chapter 5. Four endeavours [1,2,3,5], reproduced as Appendices 5-A, 5-B, 5-C, and 5-D, respectively, form a systematic investigation of the manner in which additional information may be incorporated in iterative algorithms to retrieve the phases of radiated wavefields. Appendix 5-E examines an aspect of (conventional, remote probing) computed tomography, namely the effects of possessing only discrete samples of individual projections [4].

Inverse scattering algorithms *per se* are treated in Chapter 6. Mathematical as well as numerical aspects of each technique are considered, and limitations likely to be experienced in practice are outlined. Appendix 6-A treats a problem in remote sensing computed tomography [7]. Appendices 6-B and 6-C are preprints of papers are specialised invocations of the iterative Newton-Kantorovich algorithm for solving, to arbitrary accuracy, two specific inverse problems [10,11].

Finally, Chapter 7 assesses the significance of this thesis and suggests directions for further research.

References are identified in the text in the form "[Name date]" or "Name [date]" depending on the context, with the full citation list following Chapter 7. References to books generally specify the pertinent chapter or section but, in order to maintain consistency with different editions, not specific pages.

Computer programs to implement aspects of the work contained in this thesis were written by the author in both FORTRAN and Pascal on the VAX 11-750 and MicroVAX-II facilities of the Department of Electrical and Electronic Engineering, University of Canterbury. The typesetting of this thesis has been performed with VAX implementations of TEX and LATEX, written respectively by D. E. Knuth, Stanford University, and L. Lamport, Digital Equipment Corporation, and printed on a Hewlett-Packard LaserJet Series II laser printer with the aid of software by N. H. Beebe, University of Utah.

During the course of my studies, the following papers have been published, submitted, or presented at conferences :

[1] R. H. T. Bates and D. G. H. Tan. Fourier phase retrieval when the image is complex. In A. J. Devaney and R. H. T. Bates, editors, *Inverse Optics II*, pages 54–59, Proceedings SPIE volume 558, August 1985.

[2] R. H. T. Bates and D. G. H. Tan. Towards reconstructing phases of inverse-scattering signals. *Journal of the Optical Society of America A*, 2(11):2013–2018, November 1985.

[3] P. H. Gardenier, C. A. Lim, D. G. H. Tan, and R. H. T. Bates. Aperture distribution phase from single radiation pattern measurement via Gerchberg-Saxton algorithm. *Electronics Letters*, 22(2):113–115, 16 January 1986.

[4] D. G. H. Tan, J. X. Qu, and R. H. T. Bates. Image reconstruction from projections: VIII: Effects of finite resolution and sampling of individual projections.

Optik, 73(1):25–29, April 1986.

[5] R. H. T. Bates and D. G. H. Tan. Phase retrieval for electron diffraction patterns suggesting orientational order without translational symmetry. *Optik*, 73(2):69–73, May 1986.

[6] D. G. H. Tan and R. H. T. Bates. Generalised approach to inverse diffraction. Presented at New Zealand Mathematics Colloquium, Christchurch, New Zealand, 19–21 May, 1986. Abstract in *Mathematical Chronicle*, 15:69, December 1986 (Mathematical Chronicle Committee, Mathematics and Statistics Department, University of Auckland, New Zealand).

[7] D. G. H. Tan, J. X. Qu, K. L. Garden, and R. H. T. Bates. Allowing for variable resolution and constant attenuation in SPECT. *IEE Proceedings Part A*, 134(2):136–142, February 1987.

[8] D. J. N. Wall, D. G. H. Tan, and T. S. Yeo. Inverse scattering and the null field method. Presented at Applied Mathematics Conference, Wairakei, N. Z., 8–12 February, 1987. Abstract available from Australian Mathematical Society, Applied Mathematics Division.

[9] D. G. H. Tan and R. H. T. Bates. Generalised volume source formulation of diffraction and a sequence of approximations. To appear in *Wave Motion.*

[10] D. G. H. Tan and D. J. N. Wall. Newton-Kantorovich method applied to inverse scattering for an exterior Helmholtz problem. Submitted (January 1988) to *Inverse Problems.*

[11] D. G. H. Tan, R. D. Murch, and R. H. T. Bates. Algorithmic implementation of global solution to scalar inverse scattering problem. Submitted (February 1988) to *Inverse Problems.*

Contents

Chapter 1

INTRODUCTION

1.1 PREAMBLE

This thesis is concerned with the theory of procedures for locating, and determining the structure of, distant objects from non-invasive or remote measurements. This is often called the "imaging problem" [McKinnon 1980] and falls into the realm of "inverse problems" [Santosa *et al.* 1984,Baltes 1978,Baltes 1980].

The need for imaging arises in many diverse disciplines such as radar, sonar, non-destructive testing, X-ray crystallography, electron microscopy, geophysical exploration and remote sensing of the earth [Sabatier 1978], [Boerner *et al.* 1981], [McLaughlin 1984a], [Bond and Reynolds 1987]. There are many medical imaging modalities : X-ray computed tomography (CT), single photon emission CT (SPECT), positron emission CT (PET), ultrasonic B-scans, nuclear magnetic resonance (NMR) imaging [Wells 1987,Malacara 1985], and electrical impedance tomography [Brown *et al.* 1987] are now common diagnostic tools. Motivation for studying such problems is given in Section 1.2.

Literature on imaging techniques is widespread. Because techniques have been developed independently in many disciplines, different terminology is often applied to essentially similar concepts. One of the aims of this thesis is to unify existing methods within a framework which enables comparison of the various schemes. Section 1.3 introduces the terminology for such a framework and defines the general direct and inverse problems. Whereas an inverse problem typically involves determining causes from effects, a direct problem entails predicting effects from causes. Direct problems are, on the whole, far better understood than inverse problems.

In Section 1.4, three tasks involved in any treatment of an inverse problem are identified. These are : forming a model of the physics of the situation in question, examining the direct problem, and finally devising a procedure to determine the parameters characterising the object of interest.

Considerations that follow from the desire to implement inversion algorithms numerically are intimated in Section 1.5. The relevance of the ill-posed nature of inverse problems is also discussed.

A prime illustration of the motivation presented in Section 1.2 is the problem of echo-location [Kock 1973,Wells 1977]. Section 1.6 poses the echo-location problem and outlines its solution. The organisation of Section 1.6 demonstrates the division of the inverse problem into the three tasks mentioned above. Echo-location is the prototype of the situations considered here and is significant in that a solution to the

problem exists, is conceptually simple, and is capable of producing meaningful results in practice.

Echo-location relies on many idealizations of the real world. In various situations of scientific and technical importance, these idealizations can prove inappropriate, as is discussed in Section 1.7, which also indicates some features that, if accounted for in a more sophisticated model, would significantly improve our ability to form faithful images of remote objects. This serves to establish a context in which the remainder of this thesis should be understood.

1.2 MOTIVATION

All our knowledge of the physical world is obtained through our senses (sight, hearing, touch, smell, and taste). The physical characteristics of our environment can often be determined from direct observation or measurement with the unaided senses. For example, objects within visible range can be located and identified. Oncoming traffic can be detected by listening. The surface of an object within physical reach can be touched to infer texture or temperature. The human sense of smell is considerably less developed although other animals rely extensively on smell to detect predators and prey. Tasting a substance usually has the undesirable (from the point of view of this thesis) effect of drastically changing its physical constitution.

The interest here is with the theory of techniques for forming images of the nature of objects, without sensibly altering them, when direct measurements by the unaided senses are inadequate. For example, when the sense involved is sight, an object may be occluded, as internal organs are by flesh. Alternatively, there may be detail too small to be discerned by the naked eye. In such circumstances, imaging must be performed with the use of indirect or remote measurements.

Remote imaging instruments and techniques have been with us for several centuries, and it is convenient to think of them as belonging to three "generations", as defined in the following paragraphs. Within each generation, a further distinction can be made between "remote probing" and "remote sensing" [Bates *et al.* 1983]. Often, the object being viewed is actively stimulated by the observer, in which case the imaging procedure is termed "remote probing". If the observer merely passively receives radiation emitted from the object, the imaging procedure is termed "remote sensing".

First generation remote imaging techniques are those that enhance the sensitivity of the natural senses. Instruments such as astronomical telescopes and microscopes enable viewing of objects too small or too far away to be seen by the naked eye. The use of a telescope is an example of remote sensing. The object being viewed (a star or other celestial body) emits radiation (visible light) which is captured by the telescope. If a person operating a microscope deliberately controls the light illuminating the specimen, the resulting image can be thought of as formed by remote probing. By employing such instruments, visible light (or, equivalently, audible sound) can be used to examine surfaces of objects and transparent objects. The class of objects that can be viewed this way is, however, restricted. The interior of an object remains inaccessible if light or sound cannot penetrate beneath its surface.

Second generation imaging techniques employ physical processes that can penetrate where the natural senses cannot. These processes are here called emanations

(defined formally in Section 1.3), and include electromagnetic radiation outside the visible light band, and acoustic waves beyond limits of audible sound. Because the natural senses are insensitive to general emanations, transducers are needed to convert them to a readily interpretable form. Often this form is a display on some two-dimensional surface, such as a CRT monitor. Second generation techniques are typified by the resultant images being of one dimension (at least) less than the object of interest. An image formed by such a procedure cannot, even in principle, be a faithful representation of the spatial distribution of the object, but it can nevertheless be useful. Perhaps the most effective application of a second generation imaging procedure is in medicine. X-rays passing through a human body can be detected by a fluorescent screen or photographic film. The image formed on the recording surface, which is sometimes called a shadowgram, is the line-integral (along straight rays in the direction of the probing X-ray beam) of the X-ray attenuation coefficient of the body. A shadowgram is, thus, a two-dimensional distribution whose intensity in a particular area contains contributions from a three-dimensional volume of the body. Trained clinicians routinely use shadowgrams as a diagnostic tool. A certain amount of ambiguity remains when information is presented in this way. Features in one plane of the object are not immediately distinguishable from features in another. Consequently, a shadowgram is not a faithful representation of the X-ray attenuation distribution of a body but merely a projection of it.

Third generation imaging techniques also employ the general emanations used by second generation techniques. The images formed, however, are of the same dimension as the object, and are (ideally) directly proportional to the distribution of whatever material property of the body is being imaged. The reconstruction corresponds to a region of the object of exactly the same dimension. So, whereas a shadowgram contains the superposition of many planes, a two-dimensional computed tomographic reconstruction represents a single planar cross-section through a body.

A large number of diverse scientific and technological applications depend on, or benefit from, the ability to perform remote imaging. There is a tendency for workers to remain ignorant of developments in applications with which they are not intimately involved. This is particularly unfortunate where imaging is involved because principles invoked in one application often have analogies in another. This has prompted this thesis to have the following aims.

First, existing imaging methods are extensively reviewed. The literature in specific fields is comprehensively referenced, permitting readers to pursue their own particular interests, but also indicating related work in other applications. More importantly, such methods are presented as viewed within a systematic interpretation of inverse scattering methods. Different models and techniques exhibit a variety of sophistication, ranging from very approximate (and easy to implement) to theoretically exact (and hard to implement). The approach taken in this thesis enables limitations of various theories to be identified. Furthermore, such an approach permits identification of areas where theory is lacking. The original work reported in this thesis reflects the author's research into some of these areas.

1.3 DEFINITION OF GENERAL DIRECT AND INVERSE PROBLEMS

Following Bates *et al.* [1983], the following terms are defined :

- The *experimenter* is anyone solving or wishing to determine the material properties of an object.
- The *body* represents any object whose material properties are to be determined.
- The *emanations* represent any physical process used to study the body from outside.
- The *generalised constitutive parameter* represents the spatial distribution of whatever material properties of the body the experimenter wishes to reconstruct.
- The *system* represents the combination of apparatus, measurement scheme, and emanations employed by the experimenter.
- A *neighbourhood* is the smallest spatial element (volume or cross-section) of a body, that can be resolved by the system.
- A *clean image* of the generalised constitutive parameter is one for which the value reconstructed at any spatial point is (ideally) uncontaminated by values of the true parameter outside the neighbourhood of the point.

The general problems considered in this thesis are :

the general direct problem : given the generalised constitutive parameter of the body, predict the behaviour of emanations that interact with it, and

the general inverse problem : given measurements, made remote from the body, of emanations that have interacted with the generalised constitutive parameter, reconstruct a clean image of the generalised constitutive parameter.

The general inverse problem is usefully subdivided into :

the general inverse source problem : in which the experimenter has no control over the source of the emanations. The generalised constitutive parameter to be reconstructed is the distribution of this source. This is also known as the remote sensing problem.

the general inverse scattering problem : in which the experimenter controls the source of the emanations. The generalised constitutive parameter to be reconstructed is the distribution of the material property that interacts with the emanations. This is also known as the remote probing problem.

1.4 APPROACH ADOPTED IN THIS THESIS

In Section 1.3, the general direct and inverse problems are defined descriptively. Herein is presented a general mathematical framework from which these problems can be attacked. The treatment of an inverse problem can be partitioned into three major tasks [Santosa *et al.* 1984]. The first, which is addressed in Chapter 2, is to form a model of the physics describing the situation. This involves deciding what species of emanations are of interest, what physical parameters characterise the generalised constitutive parameter, and how the two quantities interact. The second is to examine the direct problem, and occupies Chapters 3 and 4. The benefits of this are two-fold. First, the model can be verified. Typically, this is done by solving the direct problem for some test object. Theoretically predicted measurements are then compared with experiment. A thorough understanding of the direct problem also facilitates determination of what parameters can, in fact, be inferred from inverse problem data. This is discussed further in Section 1.5. The third task is to devise procedures to infer the parameters of the body from remote measurements; this represents the actual solution of the inverse problem, and is dealt with in Chapters 5 and 6.

It is hardly ever the case with significant inverse problems that the three tasks mentioned in the previous paragraph can be approached in a sequential manner. The nature of research is to advance, with the benefit of hindsight, by gradual refinement of previous efforts. Nevertheless, the peculiarities of various approaches to inverse problems seem best presented by organising them according to their respective approaches to the three tasks mentioned above, rather than adhering to strict chronological order.

The complexity of the model employed is crucial to the usefulness of any image reconstruction procedure. Simple models lead to relatively simple algorithms. One of the simplest is the echo-location model described in Section 1.6. Despite many idealizations, the success of radar, sonar and ultrasonic B-scan imaging are convincing examples that such simple models can often yield satisfactory results in practice.

There are, however, many applications where simple models are insufficient. Two approaches may then be taken. The first is to increase the sophistication of the model. This can be done by increasing the number of variables describing the generalised constitutive parameter and the emanations or, perhaps more importantly, describing more accurately the interaction between them. As might be expected, the benefit of using a more accurate physical model is offset by the need to use more complicated, if not impractical, algorithms. The second approach is to treat deviations from a simple model as perturbations. One may then attempt to compensate for such perturbations by relatively simple algorithms. In the real world, a compromise must be reached. The model is made as complicated as possible while still permitting a solution to be obtained, subject to other constraints (e.g. limited computer time, cost of obtaining measurements). In this thesis, emphasis is placed on models permitting reconstruction algorithms which are practicable, given current technology or imminent technological advances.

The most general species of emanations must be represented by vector fields satisfying nonlinear equations of motion. In the present state of understanding of inverse theory it is profitless to start from such an unconstrained basis. The following is the most general model considered in this thesis. The body is contained in a region Υ_s (refer to Figure 1.1). The generalised constitutive parameter is denoted by χ. The

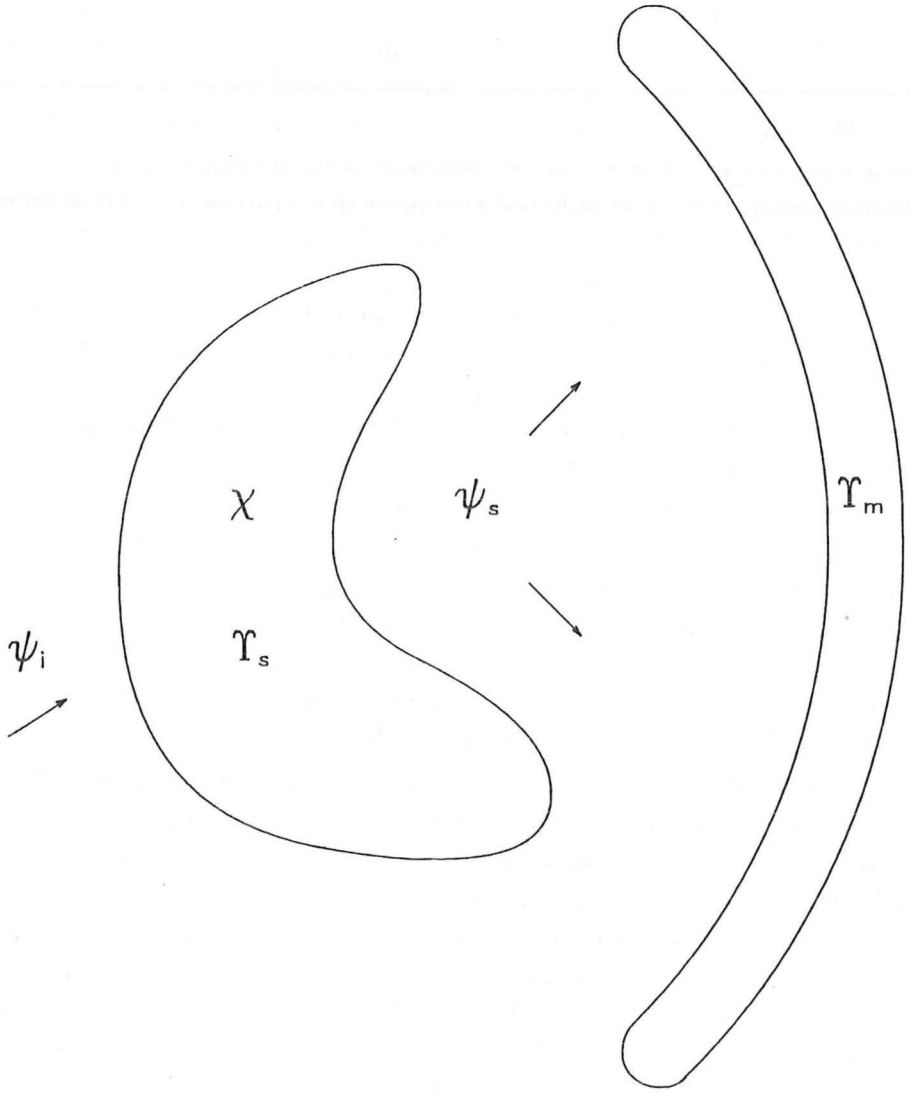

Figure 1.1: General geometry for direct and inverse problems. The generalised constitutive parameter χ is contained in a region Υ_s. The total emanation ψ is the sum of two fields ψ_i and ψ_s. Measurements for the inverse problem are made in the region Υ_m.

emanations are represented mathematically by a scalar field ψ which, at each point in space and time (or, equivalently, at all temporal frequencies), assigns a specific value to the emanations. This value is in general a complex number, and is often a single component of a vector field, although many applications can be characterised by an intrinsically scalar field. It is convenient to partition ψ into the sum of two fields

$$\psi = \psi_i + \psi_s, \tag{1.1}$$

where ψ_i is (ideally) completely under the control of the experimenter, and ψ_s is the difference between ψ and ψ_i. Often, ψ_i represents some initial or incident field, and ψ_s the secondary or scattered field. The interaction between the emanations and the generalised constitutive parameter is described by an operator equation, here indicated symbolically by

$$\psi = \Omega\{\psi_i, \chi\}. \tag{1.2}$$

The direct problem can be considered as finding the result of the mapping (1.2) when Ω, χ, and ψ_i are all given. Since ψ_i is given, the unknown in the direct problem is effectively ψ_s. It may be required to determine ψ_s either inside or outside Υ_s. The mapping Ω is often linear with respect to ψ_i, i.e.

$$\Omega\{a\psi_i + b\phi_i, \chi\} = a\Omega\{\psi_i, \chi\} + b\Omega\{\phi_i, \chi\}, \tag{1.3}$$

for any complex numbers, a and b, and any incident fields, ψ_i and ϕ_i. It is only when extensive simplifications are made that Ω is also linear with respect to χ.

In the inverse problem, ψ is measured in the region Υ_m, remote from Υ_s. The reconstruction of the generalised constitutive parameter can be expressed in terms of the inverse to the direct problem operator, i.e.

$$\chi = \Omega^{-1}\{\psi_i, \psi\}. \tag{1.4}$$

Note that (1.4) cannot be implemented directly unless ψ is known within Υ_s, which is rarely the case. In fact, examination of (1.2) shows that ψ depends on χ, making (1.4) an implicit, nonlinear, equation for the unknown χ. Thus, the inverse problem is generally more difficult than the direct problem.

1.5 DISCRETE VERSUS CONTINUOUS REPRESENTATION OF FUNCTIONS AND ILL-POSEDNESS

The relationships between χ and ψ as defined by Ω and its inverse are normally so intricate that it is necessary to resort to numerical computation. While the description of the mathematical physics defining Ω is usually best performed by treating ψ and χ as functions of continuous variables, any computer implementation can only involve a finite number of quantities, whose values are themselves quantised. Typically, the quantisation is fine enough to be neglected in the sense that any finite representation of χ can be considered [Harrington 1968, Chapter 1] an approximation in terms of

known basis functions, Ξ_m say, whose coefficients χ_m are essentially arbitrary complex numbers, i.e.

$$\chi = \sum_{m=1}^{M} \chi_m \Xi_m . \tag{1.5}$$

The finiteness of the summation is emphasised by the upper limit M. The summation (1.5) permits a great deal of flexibility in the choice of basis functions. Usually the minimum stipulation is that, in the limit as M tends to infinity, the error (measured in whatever manner is deemed appropriate) in the approximation decreases to zero.

The degree of approximation associated with (1.5) depends on the nature of the functions being approximated, the basis functions Ξ_m, and also M. The analytical results that exist for certain situations [Cheney 1969,Rivlin 1969] tend to be of little practical significance, because the effect of a finite approximation on an inversion algorithm usually eludes analysis when the direct mapping is nonlinear. While it would be desirable to have analytical means by which to estimate the magnitude of M required to achieve a specific accuracy, the former is invariably chosen by computational experiment, that is, by the unimaginative procedure of solving a progression of numerical problems for successively larger values of M until the true solution is adjudged to have been attained (to within the desired accuracy) [Bates 1975a].

The complexity of inversion algorithms is such that the computational effort of manipulating inverse problem data is often more demanding than storage limitations. As M is increased, the predominant limitation of computer resources is often the amount of computation time required to solve the discretised problem which grows far more rapidly than storage requirements (see, for example [Johnson and Tracy 1983], [Tracy and Johnson 1983], [Johnson *et al.* 1984]).

The nature of many direct problems is such that their corresponding inverse problems are "ill-posed" [Deuflhard and Hairer 1983,Hämmerlin and Hoffmann 1983], [Sabatier 1983], [Colton 1984], [Cannon and Hornung 1986], [Craig and Brown 1986]. One consequence of ill-posedness is that, although the error associated with the finiteness of M in (1.5) decreases with increasing M, other sources of error (such as noise and finiteness of quantisation) assume a greater, and sometimes dominant, significance [Nashed 1981]. This prevents solutions of arbitrary accuracy being attained. In general there is an optimum value $\tilde{M}$ for M which provides a compromise between reducing the approximation error and amplifying the round-off error (see Figure 1.2) [Natterer 1983]. The best choice for $\tilde{M}$ depends on circumstances beyond the influence of the direct problem itself, such as the level of noise present on the data, and judgement can only be made with the aid of *a priori* information.

A more systematic approach to overcoming ill-posedness is to introduce *a priori* information in such a way as to replace the original ill-posed problem by a well-posed problem. Typically, this can be done by requiring the solution to the well-posed problem to satisfy closely, but not exactly, the original problem while possessing other desirable properties such as smoothness or positivity [Backus and Gilbert 1970], [Deschamps and Cabayan 1972].

The technique known as Tihonov (Tikhonov) [1963a,b] regularisation is an implementation of the above ideas [Craig and Brown 1986, Chapter 5]. The original ill-posed problem is replaced by an optimisation problem, to be solved by conventional means [Fletcher 1980], [Fletcher 1981], [Gill *et al.* 1981].

Solution of the optimisation problem is only likely to be effective if the original

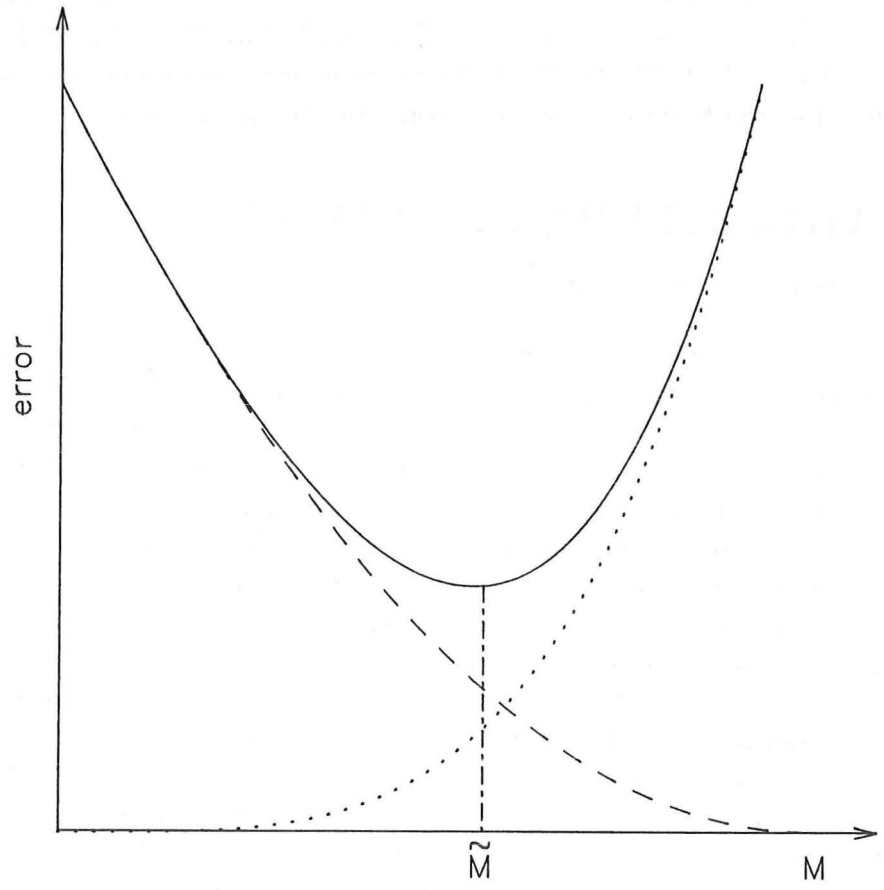

Figure 1.2: The finite number M of basis terms leads to approximation error (the dashed curve) which decreases with increasing M. Ill-posedness causes other errors to increase with M (the dotted curve). The combined effect is to produce an optimum $\tilde{M}$ for which the combined error (the solid curve) is least.

ill-posed problem has at least a reasonable conceptual (i.e. feasible in the ideal situation of no noise) solution. The essential concepts underlying inversion algorithms can be understood without explicit consideration of ill-posedness. Of course, any numerical implementation should not disregard ill-posedness but regularisation techniques to overcome ill-posedness can often be incorporated simply into the numerical scheme (see, for example, Appendix 6-B). For this reason, remedies for ill-posedness are not explicitly considered in the body of this thesis (only in the appendices).

1.6 AN ARCHETYPAL EXAMPLE: ECHO-LOCATION

Suppose a homogeneous medium exists with constant speed of propagation and constant attenuation coefficient both known. It suffices to consider this medium to be confined to a plane. The arguments of this section and the next easily generalise to three (or more) dimensions. A pulse is transmitted from Tx (see Figure 1.3a). If no scatterers are present the pulse radiates, unperturbed, uniformly in all directions. Assume that there are small scatterers embedded in the medium, whose locations and amplitudes are to be determined. The scatterers must be small enough to be considered isolated point scatterers, as defined below. Two scatterers are deemed isolated if they are sufficiently far apart to ensure that the time taken for a pulse to travel between them is more than the duration of the pulse itself. An individual scatterer can be considered a point scatterer if any radiation impinging on it is scattered, to a good degree of approximation, uniformly in all directions. Assume further, that the scatterers are weak, first in the sense that radiation from Tx incident upon a scatterer gives rise to a single scattered wave, and second in that multiply scattered signals can be neglected. A receiver at Rx intercepts part of the scattered radiation. Thus, the signal recorded here consists of delayed (with respect to the transmitted pulse) echo pulses of varying amplitudes (see Figure 1.3b).

Since the speed of propagation is constant, the set of points consistent with a given time delay form an ellipse having Tx and Rx as foci, as shown in Figure 1.4a. The attenuation due to a scatterer at this range can be calculated and the amplitude of the scatterer deduced. In order to locate the precise position of the scatterer it is necessary to make further measurements. By moving Tx, or Rx, or both, different ellipses are produced, all intersecting only at the true positions of the scatterers (see Figure 1.4b), thereby solving this simple inverse scattering problem.

1.7 BEYOND ECHO-LOCATION AND STRAIGHT RAYS

The echo-location principle introduced in the previous section is the basis of radar, sonar, and medical B-scan imaging . Its appeal can be attributed to the simple model of the physics involved. In radar, the emanations employed are electromagnetic waves, while the latter examples employ acoustic radiation. Because it is so simple to understand and to apply, it is tempting to invoke the echo-location principle whenever one attempts to probe a distant or opaque medium (of any kind) with arbitrary wave-

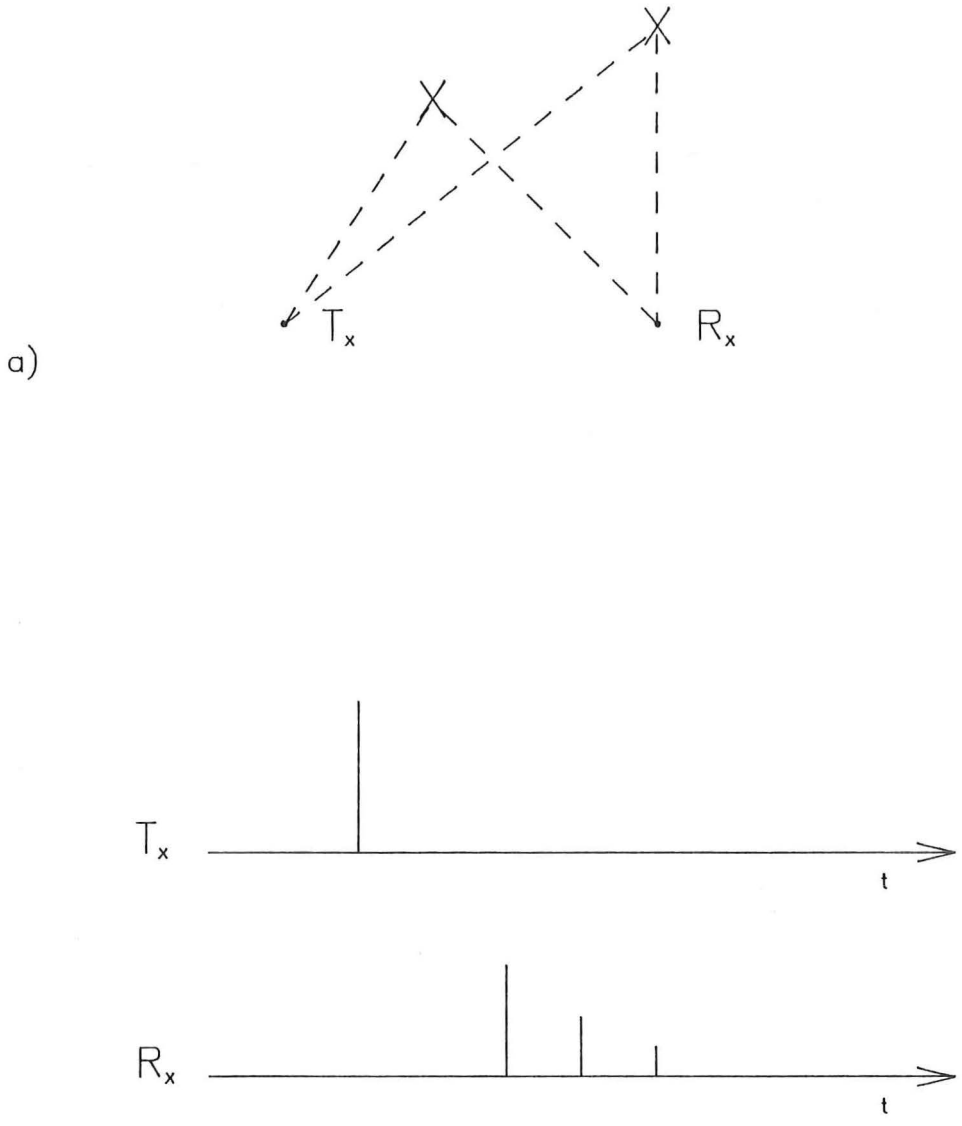

Figure 1.3: (a) Scatterers are situated at points X. Tx transmits an incident pulse. Rx receives the scattered field which has travelled along the dashed paths. (b) The transmitted and received time waveforms.

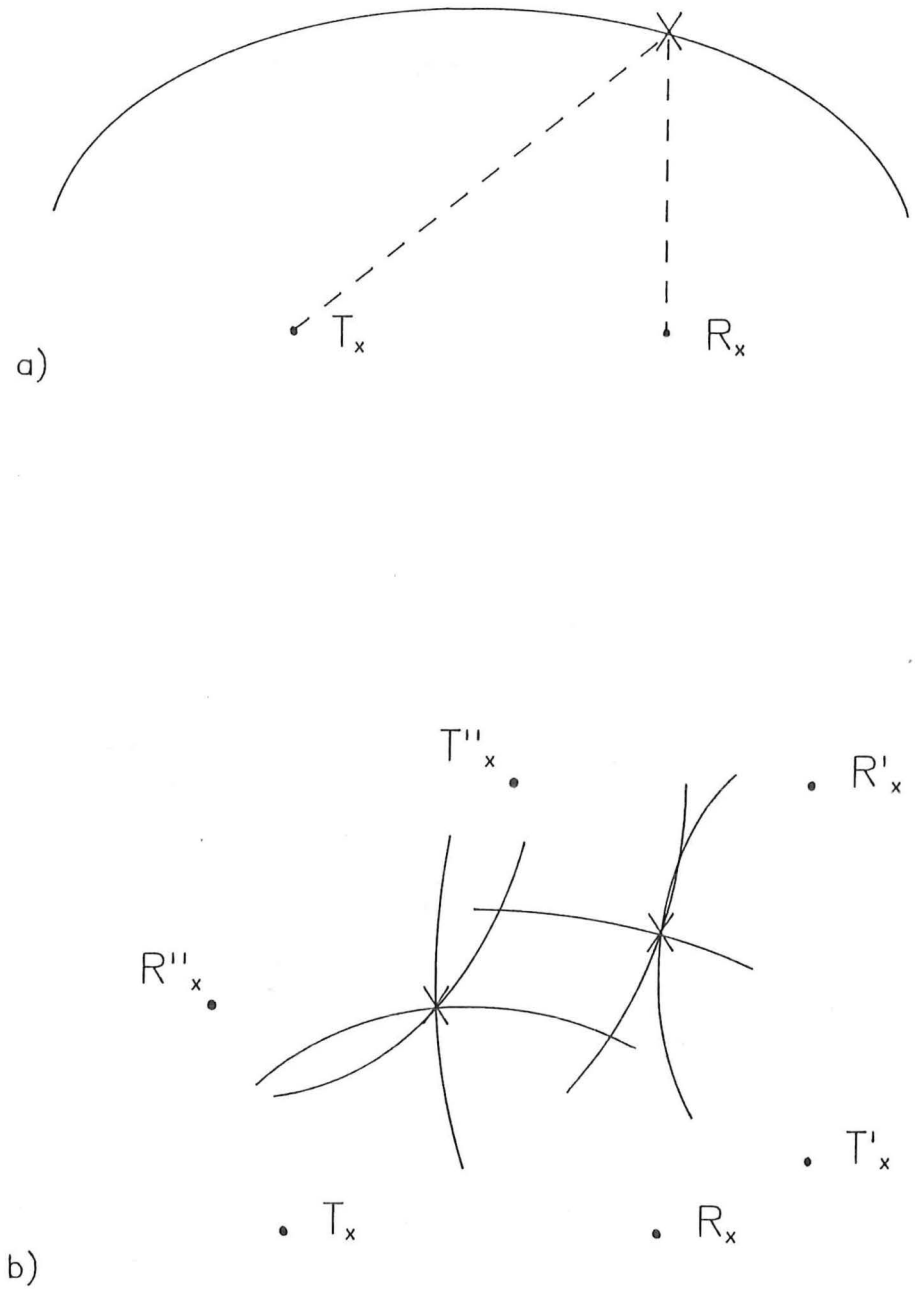

Figure 1.4: (a) All travel times from Tx to points on the ellipse (solid line), and then to Rx, are the same. (b) Ellipses generated by different Tx–Rx pairs all intersect only at the true positions of the scatterers.

like emanations. However, caution must be exercised. For the echo-location principle to be useful, the emanations and their interaction with the generalised constitutive parameter must satisfy various constraints. In situations where these constraints are significantly violated, and the behaviour of the emanations is no longer simple, it is necessary to increase the sophistication of the model in order to maintain an adequate description of the interaction. It is desirable to resist the introduction of more complications than are absolutely necessary, both for conceptual and practical considerations. In the remainder of this section, the assumptions underlying the echo-location principle are identified. Departures from these assumptions likely to be significant in practice, and possible ways of compensating for them, are outlined.

First, as is indicated in Section 1.6, it must be assumed that the medium can support wave motion, and that the waves travel with a constant speed of propagation. The wave motion can then often be adequately described in terms of straight rays and simple laws of reflection, enabling the problem to be solved geometrically. This assumption is implicit in all applications of radar [Kock 1973,Skolnik 1962]. Such a description is equivalent to requiring the radiation to satisfy Fermat's principle of least action in a homogeneous medium, and is encompassed in the theory of geometrical optics [Kline and Kay 1965]. The propagation speed is governed for electromagnetic and acoustic phenomena, respectively, by the permittivity and permeability, and by the density and compressibility, of the medium [Jones 1964]. Variations in these parameters cause the propagation speed to vary. If the variations are small enough, it may be possible to describe the wave motion with curved rays [Cornbleet 1983]. Time delay, however, is no longer directly proportional to range. Variations in attenuation coefficient have a similar effect on amplitude estimation. In general, propagation speed variations cause reflection, refraction, and diffraction of the wave motion [Jones 1964]. Depending on the extent of the variations, and on the application, these effects may render echo-location totally or partially ineffective.

Secondly, it is assumed in Section 1.6 that each scattering object is small enough to be associated with a single point in the medium. This depends on the duration of the pulse employed. If two scatterers are too close to be considered isolated, then any received echoes they generate overlap. If the received echoes overlap sufficiently, the scatterers become virtually indistinguishable and are said to be unresolvable. An intuitive solution to this difficulty is to reduce the duration of the pulse. However, in order to radiate a finite amount of energy, the duration of the pulse must remain above some finite level. Furthermore, the maximum size of scatterer that can be considered a point is directly proportional to the pulse duration. By reducing the duration of the pulse, the radiation scattered from an object ultimately ceases to be isotropic and acquires significant directivity. Thus there is an unavoidable compromise here between reducing the pulse duration, which improves resolution, and maintaining the validity of the assumption of point scatterers. The geometrical theory of diffraction [Keller 1962,James 1976], the theory of physical optics [Bouwkamp 1954,Bates 1974], and the physical theory of diffraction [Ufimtsev 1971] are three approaches to describing scattering by such extended objects (those that are not point scatterers). Moffat *et al.* [1981] use techniques such as the singularity expansion method [Baum 1976] to describe the scattered signals, and to infer the structure of the scatterer. All these approaches represent considerable departures from the simple concepts underlying the echo-location principle.

Another assumption made in Section 1.6 is that multiple scattering is negligible. If this is invalid, received signals contain extraneous returns, so that false targets are predicted when the signals are interpreted in basic echo-location terms. Because scattered signals spread out and are attenuated as they propagate away from a scatterer, the strength of multiply scattered signals depends on the distance between scatterers. As the separation between individual scatterers decreases, the significance of multiply scattered signals tends to increase.

Table 1.1 lists a selection of emanations, constitutive parameters with which they interact, transducers which transmit and receive them, and applications in which they may potentially be used for imaging. Notably, the schemes which are currently most successful are those based on the principles underlying echo-location, which is why echo-location has been introduced in this opening chapter. For the imaging methods listed in Table 1.1, that are predicated on physical models more sophisticated than those invoked for echo-location, many of the difficulties associated with them can be attributed to the incompatibility of each model with one or more of the assumptions underlying the echo-location principle. It is with overcoming these difficulties that this thesis is concerned.

Emanations	Constitutive parameters	Transducers	Imaging Applications
X-rays	X-ray attenuation coefficients (elastic and Compton scattering besides true absorption)	X-ray sources; scintillation detectors	Diagnostic radiology (DR); non-destructive testing (NDT); package and luggage surveillance
Gamma-rays from radio-isotope labelled ingested or injected substances	Concentration of radio-labelled substance	Scintillation counters	DR
Electron (wave) beams	Schrödinger potential distribution	Electron guns; film; photomultiplier/TV	Microscopy; tomography
Acoustic and elastic waves	Density; compressibility; refractive index; attenuation	Any electromechanical transduction device	DR; NDT; geological prospecting (GP); underwater sonar
Low-frequency electric currents	Electrical conductivity distribution	Electrodes	Crude imaging of blood vessels; GP; NDT
Magnetic fields	Distribution of nuclear spins; blood flow	(Electro)magnets; RF coils; magnetometers	Magnetic resonance imaging (MRI); magnetohydrodynamic imaging
Radio frequency (RF) and microwave fields	Electron spins; permittivity and conductivity distributions	Capacitors; coils; antennas	MRI; DR; GP; radar; determination of atmospheric material and electron densities; microwave hyperthermia imaging
Spatially incoherent electromagnetic radiation from microwaves to X-rays	Celestial brightness distributions; volume temperature distributions	Antennas; X-ray, ultraviolet, optical, infrared, radio telescopes	Astronomical imaging; mapping temperature distributions inside human and other bodies

Table 1.1: Emanations and generalised constitutive parameters

Chapter 2

MATHEMATICAL PHYSICS OF SCATTERING

2.1 INTRODUCTION

This chapter is concerned with the first task involved in examining an inverse problem, namely, forming a mathematical model of the physics describing the behaviour of emanations that interact with whatever objects occupy the region of interest. In Section 2.2 is presented a description of the most general time-domain behaviour of the emanations considered in this thesis. This description is defined by what is termed the underlying time-domain equation, and is a scalar linear partial differential equation in which the constitutive parameters characterising the afore-mentioned objects appear as coefficients. The equations employed in standard models of the behaviour of some of the emanations listed in Table 1.1 are manipulated and shown to be special cases of the underlying time-domain equation. It is considered instructive to include such details in order to enable identification of the assumptions required in the physical models, to reduce the equations to the underlying form. Thus, circumstances in which such a model is invalid can be determined. In Section 2.3, the equivalent frequency-domain formulation of the underlying time-domain equation is presented.

In Section 2.4 are presented certain transformations to convert the general formulations of the first two sections to forms more suitable for examining direct and inverse problems. First, the underlying frequency-domain equation is subjected to a transformation converting it to what is termed the canonical form, involving a generalised constitutive parameter. The corresponding time-domain transformation and its limitations are discussed.

There are two forms of generalised constitutive parameter widely considered in inverse scattering literature. One of these is that due to a variable refractive index, leading in the frequency domain to the Helmholtz equation with inhomogeneous wavenumber. The other is that due to a quantum mechanical scattering potential, leading to the Schrödinger equation. These two forms lead to quite different approaches to the respective inverse problems. The second transformation indicated in Section 2.4, valid for situations in which the constitutive parameter due to a variable refractive index depends on only one spatial variable, can be invoked to convert the inhomogeneous Helmholtz equation into the Schrödinger equation. In such circumstances, techniques originally devised in quantum mechanical contexts can thus be applied to problems

concerned with wave scattering from regions of variable refractive index.

In addition to the equations of Sections 2.2 through 2.4, it is necessary to introduce supplementary equations in order to completely describe the behaviour of emanations. Typical supplementary equations are collected in Section 2.5.

In Section 2.6, the relative merits of the various time- and frequency- domain formulations are assessed. It is concluded that, in order to proceed with a unified treatment of all the phenomena considered in this chapter, the frequency-domain approaches are more suitable. This is in accordance with the majority of inverse scattering literature.

The concept of temporal frequency can be extended to spatial frequency. Section 2.7 introduces the multi-dimensional Fourier transform and itemises some standard identities [Bracewell 1978], [Bates and McDonnell 1986, Chapter 2] that are usefully invoked in later chapters.

2.2 TIME-DOMAIN EQUATIONS

We take the emanations to be adequately represented in the time domain by a complex-valued scalar field $\psi(\vec{x}\,|\,t)$ that exists for all time, $t \in R$, and all points in space, $\vec{x} \in \Upsilon$. The temporal variable is preceded by '|' to avoid ambiguity with the frequency-domain field $\psi(\vec{x}, k)$ introduced in Section 2.4. No restriction is placed on the dimension K of the vector $\vec{x}$, whose components are denoted $(x_1, \ldots, x_K)$, although in two spatial dimensions it is more convenient to write (x, y). The underlying equation, describing the most general behaviour of the emanations considered in this thesis, is

$$\left[\mathcal{L}_1(\vec{x}\,|\,\partial_t)\,\nabla\cdot\left(\gamma^{-1}\nabla\right) - \mathcal{L}_0(\vec{x}\,|\,\partial_t)\right]\psi(\vec{x}\,|\,t) = -\Sigma(\vec{x}\,|\,t), \quad \vec{x}\in\Upsilon, \quad t\in R, \tag{2.1}$$

where ∂_t denotes partial differentiation with respect to time, and ∇ is the spatial gradient operator. The operators $\mathcal{L}_0$ and $\mathcal{L}_1$ are differential operators of the form

$$\mathcal{L}_1(\vec{x}\,|\,\partial/\partial t) = \sum_{n=0}^{N_1} \alpha_n(\vec{x})\,\partial^n/\partial t^n, \tag{2.2}$$

$$\mathcal{L}_0(\vec{x}\,|\,\partial/\partial t) = \sum_{n=0}^{N_0} \beta_n(\vec{x})\,\partial^n/\partial t^n, \tag{2.3}$$

where the α_n, β_n, and γ are constitutive parameters of the region Υ. It is assumed that these parameters are time-invariant, at least for the duration needed to gather whatever data are required. For inverse source problems, as defined in Section 1.3, Σ is also a constitutive parameter of the scattering region and must also satisfy the condition of the preceeding sentence. For inverse scattering problems, however, Σ represents the probing emanations that are under the control of the experimenter, and may vary arbitrarily in space and time, provided this variation is completely known to the experimenter. Note that the spatial derivative operator is in self-adjoint form $\nabla\cdot(\gamma^{-1}\nabla)$ [Copson 1975, Chapter 5] and that γ is assumed to be strictly positive. For almost all of the physical situations that are adequately described by the linear equation (2.1), the integers N_0 and N_1 are either 0, 1, or 2. When the equivalent frequency-domain formulation is introduced in Section 2.3, it is seen that N_0 and N_1

can in principle be any non-negative integers. Outside Υ_s the constitutive parameters assume free-space values, here defined to be

$$\begin{aligned} \gamma(\vec{x}) &= 1, \\ \alpha_0(\vec{x}) &= 1, \quad \text{all other } \alpha_n(\vec{x}) = 0, \\ \beta_2(\vec{x}) &= 1, \quad \text{all other } \beta_n(\vec{x}) = 0, \\ \Sigma(\vec{x}) &= 0, \quad \text{whenever } \vec{x} \in \Upsilon \setminus \Upsilon_s, \end{aligned} \tag{2.4}$$

where the equation for Σ only applies to inverse source problems. Note that the notation $\Upsilon \setminus \Upsilon_s$ denotes the complement of Υ_s in Υ. Thus, the free-space equation is merely the wave equation for a homogeneous medium whose speed of propagation is normalised to be unity,

$$\left(\nabla^2 - \partial_{tt}\right)\psi = -\Sigma. \tag{2.5}$$

It is perhaps surprising, given the simple nature of wave propagation in a homogeneous medium, that an equation of such generality as (2.1) is required. However, the simplicity of the wave equation in a homogeneous medium is deceptive. When inhomogeneities are introduced (without inhomogeneities there are no inverse problems!), it is very easy to make what are in fact unjustified generalisations of the homogeneous model, and to arrive thereby at an exceedingly inaccurate description of the behaviour of the emanations. (Hopefully, in such circumstances, careful examination of the behaviour predicted theoretically by the direct problem can be found to be inconsistent with actual measurements, so that significant errors in the model can be detected.) For this reason, the equations governing macroscopic electromagnetism, elastic wave motion, and acoustic wave motion are derived in some detail below. These derivations state precisely the assumptions and approximations involved, and demonstrate that the underlying equation (2.1) contains the minimal amount of complication required to provide a unified treatment of the inverse scattering procedures considered in this thesis.

2.2.1 Electromagnetism

For the purposes of this thesis, macroscopic electromagnetic phenomena are adequately described by Maxwell's equations [Jones 1964, Chapter 1] :

$$\nabla \cdot \vec{B} = 0, \tag{2.6a}$$

$$\nabla \cdot \vec{D} = \rho, \tag{2.6b}$$

$$\nabla \wedge \vec{H} = \vec{J} + \dot{\vec{D}}, \tag{2.6c}$$

$$\nabla \wedge \vec{E} = -\dot{\vec{B}}, \tag{2.6d}$$

where $\vec{E}$ is the electric field vector, $\vec{H}$ is the magnetic field vector, $\vec{D}$ is the electric flux density, $\vec{B}$ is the magnetic flux density, $\vec{J}$ is the current density, and ρ is the charge density. All bodies considered here are taken to be linear and isotropic with constitutive equations of the form [Jones 1964, Chapter 1]

$$\vec{D} = \epsilon \vec{E}, \tag{2.7a}$$

$$\vec{B} = \mu \vec{H}, \tag{2.7b}$$

$$\vec{J} = \sigma \vec{E}, \tag{2.7c}$$

where the permittivity ϵ, the permeability μ, and the conductivity σ are real, scalar functions of space characterising a particular medium. Media not considered in this thesis include those whose constitutive parameters vary sufficiently rapidly in time to contribute significantly in the time-derivative terms of Maxwell's equations, and anisotropic materials whose constitutive parameters are tensors.

Substituting (2.7b) into (2.6d) produces

$$\nabla \wedge \vec{E} = -\mu \dot{\vec{H}}. \tag{2.8}$$

Taking the curl of both sides and invoking a vector identity [Arfken 1970, Chapter 1] yields

$$\begin{aligned} \nabla(\nabla \cdot \vec{E}) - \nabla^2 \vec{E} &= \nabla \wedge \nabla \wedge \vec{E} \\ &= -\nabla \wedge (\mu \dot{\vec{H}}) \\ &= -\nabla \mu \wedge \dot{\vec{H}} - \mu \nabla \wedge \dot{\vec{H}}. \end{aligned} \tag{2.9}$$

Expanding the RHS of (2.9), with the aid of (2.8) and the time derivative of (2.6c), gives

$$\nabla(\nabla \cdot \vec{E}) - \nabla^2 \vec{E} = \nabla \mu \wedge (1/\mu) \nabla \wedge \vec{E} - \mu(\dot{\vec{J}} + \ddot{\vec{D}}). \tag{2.10}$$

By employing (2.7a) and (2.7c) and rearranging,

$$\nabla^2 \vec{E} - \mu \epsilon \ddot{\vec{E}} - \mu \sigma \dot{\vec{E}} = \nabla(\nabla \cdot \vec{E}) - (\nabla \ln \mu) \wedge (\nabla \wedge \vec{E}), \tag{2.11}$$

is easily obtained. Furthermore, substituting (2.7a) into (2.6b) results in

$$\nabla \cdot (\epsilon \vec{E}) = \rho, \tag{2.12}$$

and expanding the divergence leads to

$$\nabla \cdot \vec{E} = \rho/\epsilon - \vec{E} \cdot \nabla(\ln \epsilon), \tag{2.13}$$

so that (2.11) becomes

$$\nabla^2 \vec{E} - \mu \epsilon \ddot{\vec{E}} - \mu \sigma \dot{\vec{E}} = \nabla(\rho/\epsilon) - \nabla \left(\vec{E} \cdot \nabla(\ln \epsilon) \right) - (\nabla \ln \mu) \wedge (\nabla \wedge \vec{E}). \tag{2.14}$$

For all negligibly magnetic materials (i.e. the majority), it is sufficient to approximate μ by μ_0 = permeability of free space, so that the last term of RHS (2.14) vanishes. Now, if the functions ϵ and σ vary with only one spatial variable x, say, and if $\vec{E}$ is entirely directed (i.e. is polarized) perpendicular to the x-direction, the second term in RHS (2.14) also vanishes. Thus, the equation arising from (2.14) for this perpendicular component of $\vec{E}$ is of the form of the underlying equation,

$$\nabla^2 E - \nu^2 \ddot{E} - \beta \dot{E} = -\Sigma, \tag{2.15}$$

where

$$\nu = \sqrt{\epsilon \mu_0}, \tag{2.16}$$

$$\beta = \mu \sigma, \tag{2.17}$$

$$\Sigma = -\partial(\rho/\epsilon)/\partial x, \tag{2.18}$$

and the units are chosen so that the free space parameters ϵ_0 and μ_0 satisfy (2.4), i.e.

$$\epsilon_0 \mu_0 = 1. \tag{2.19}$$

If ϵ varies with more than one spatial variable but, nonetheless, much less rapidly than $\vec{E}$, then the second term in RHS (2.14) can still be neglected, so that

$$\nabla^2 \vec{E} - \nu^2 \ddot{\vec{E}} - \beta \dot{\vec{E}} = \nabla(\rho/\epsilon). \tag{2.20}$$

Even though (2.20) is a vector wave equation, if $\vec{E}$ is expressed in a Cartesian coordinate system, so that the unit vectors are constant in direction as well as in magnitude [Arfken 1970, Chapter 2], then each component of $\vec{E}$ satisfies an equation of the form of (2.15).

In electrostatics, all the time derivatives are identically zero. Hence, (2.6d) implies that $\vec{E}$ can be written in terms of a potential function

$$\vec{E} = -\nabla V. \tag{2.21}$$

Equation (2.12) now becomes

$$\nabla \cdot (\epsilon \nabla V) = -\rho, \tag{2.22}$$

which is an elliptic form of the underlying equation. The significance of this is further discussed in Section 2.6. Note that it is not necessary to assume μ is constant here, nor that $\ln(\epsilon)$ varies negligibly in comparison with V, as was assumed in deriving (2.15).

2.2.2 Elastic waves

The physics of elastic and acoustic waves is formulated here within the framework of continuum mechanics [Spencer 1980], which attempts to describe macroscopic phenomena that occur in typical laboratory conditions on earth. Thus, velocities are assumed small relative to the speed of light, enabling relativistic effects to be ignored. Furthermore, the "continuum particle" is not a microscopic quantity but an aggregate of atomic particles over a sufficiently large volume to justify considering quantities such as density to be continuous.

There are two alternative, but completely equivalent, approaches commonly used to describe motion of a continuum [Truesdell 1977, Chapter 2]. These are the Lagrangian (sometimes called referential or material) description and the Eulerian (or spatial) description, and are distinguished by the choice of independent variables. In the Lagrangian description, the independent variables are time, t, and the positions, $\vec{X}$, of material particles at some convenient reference time. Motion is then described by the positions of these particles at later times,

$$\vec{x} = \vec{x}(\vec{X} \,|\, t). \tag{2.23}$$

Displacement and velocity are defined, respectively, by

$$\vec{u}(\vec{X} \,|\, t) = \vec{x}(\vec{X} \,|\, t) - \vec{X}, \tag{2.24}$$

$$\vec{v}(\vec{X} \,|\, t) = \frac{\partial u}{\partial t}(\vec{X} \,|\, t). \tag{2.25}$$

In the Eulerian description, the independent variables are time and fixed points in space, $\vec{x}$. At any time t, a specific material particle occupies the point $\vec{x}$, i.e.

$$\vec{X} = \vec{X}(\vec{x} \,|\, t). \tag{2.26}$$

The convective derivative of a quantity $\phi(\vec{x} \,|\, t)$ is defined by

$$\frac{D\phi}{Dt} = \frac{\partial \phi}{\partial t} + (\vec{v} \cdot \nabla)\, \phi, \tag{2.27}$$

where $\partial\phi/\partial t$ is ordinary partial differentiation with respect to t (i.e. $\vec{x}$ is held constant). In the convective derivative, the motion of the particles is taken into account by the second term in RHS (2.27).

The law of conservation of mass stipulates that the increase in mass of a volume moving with the medium is equal to the mass flowing into the volume. In differential form, writing the density of the medium as ρ, this becomes the continuity equation [Truesdell 1977, Chapter 3] :

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \vec{v}) = 0, \tag{2.28}$$

or, making use of (2.27),

$$\frac{D\rho}{Dt} + \rho \nabla \cdot \vec{v} = 0. \tag{2.29}$$

Newton's second law of motion states that the time rate of change of linear momentum of a region is proportional to the resultant force applied to it, producing the equation of motion [Truesdell 1977, Chapter 3] :

$$\rho \frac{D\vec{v}}{Dt} = \nabla \cdot \mathcal{T} + \rho \vec{f}. \tag{2.30}$$

In (2.30), $\vec{f}$ represents the external body force per unit mass. The constitution of the medium is described by the Cauchy stress tensor $\mathcal{T}$. For linear isotropic elastic media [Spencer 1980],

$$\mathcal{T}_{ij} = \mathcal{T}_{ji} = \lambda \nabla \cdot \vec{u} \delta_{ij} + \mu \left(u_{i,j} + u_{j,i} \right), \tag{2.31}$$

where $\lambda(\vec{x})$ and $\mu(\vec{x})$ are the Lamé parameters, δ is the Kronecker delta, and the partial derivative of a scalar quantity φ with respect to the i-th spatial coordinate is denoted by

$$\varphi_{,i} = \partial \varphi / \partial x_i, \tag{2.32}$$

with the obvious generalisation to higher order derivatives.

Let $\rho_0(\vec{x})$ denote the density of the medium at rest. If there are no body forces then the medium remains at rest. Now consider only a small perturbation from the static case. The displacement and density are expressed as

$$\vec{u}(\vec{x} \,|\, t) = \epsilon \vec{U}(\vec{x} \,|\, t) \tag{2.33}$$

$$\rho(\vec{x} \,|\, t) = \rho_0(\vec{x}) + \epsilon \rho_1(\vec{x} \,|\, t), \tag{2.34}$$

$$\vec{f}(\vec{x} \,|\, t) = \epsilon \vec{f}_1(\vec{x} \,|\, t), \tag{2.35}$$

where ϵ is a small dimensionless number. Substitution of (2.33) into (2.31) and retaining only terms of first order in ϵ produces, in the notation of (2.32), and with implied summation over any repeated index,

$$\begin{aligned}(\nabla\cdot\mathcal{T})_i &\equiv \mathcal{T}_{ij,j} \\ &\approx \left(\lambda\nabla\cdot\vec{U}\right)_{,i}+\mu\left(U_{i,jj}+U_{j,ij}\right) \\ &\approx \left[(\lambda+\mu)\,\nabla\cdot\vec{U}\right]_{,i}+\mu\nabla^2U_i. \end{aligned} \tag{2.36}$$

where it has been necessary to assume that gradients of μ are negligible. Thus, using a vector identity [Arfken 1970, Chapter 1],

$$\nabla\cdot\mathcal{T} = \nabla\left[(\lambda+2\mu)\,\nabla\cdot U\right]-\mu\nabla\wedge\nabla\wedge\vec{U}. \tag{2.37}$$

The convective derivative in (2.30) is nonlinear in $\vec{v}$. Since only small amplitude wave motion is being considered, terms nonlinear in ϵ can be ignored. So with the aid of (2.37), (2.30) becomes

$$\rho_0\partial^2\vec{U}/\partial t^2 = \nabla\left[(\lambda+2\mu)\,\nabla\cdot\vec{U}\right]-\mu\nabla\wedge\nabla\wedge\vec{U}+\rho_0\vec{f}_1. \tag{2.38}$$

Thus, to first order in ϵ, perturbations of the static density $\rho_0(\vec{x})$ have no effect on the resultant wave motion. Furthermore, ρ_0 commutes with time derivatives. Now, any vector $\rho_0\vec{u}$ can be written as [Morse and Feshbach 1953, Chapter 1]

$$\rho_0\vec{U} = \nabla\psi+\nabla\wedge\vec{A}, \tag{2.39}$$

and similarly,

$$\rho_0\vec{f}_1 = \nabla F+\nabla\wedge\vec{G}, \tag{2.40}$$

Substitution of (2.39) and (2.40) into (2.38), and separation of the gradient and curl terms, respectively, results in the pair of equations

$$\ddot{\psi} = (\lambda+2\mu)\,\nabla\cdot\left[\rho_0^{-1}\left(\nabla\psi+\nabla\wedge\vec{A}\right)\right]+F, \tag{2.41}$$

$$\vec{A} = -\mu\nabla\wedge\left[\rho_0^{-1}\left(\nabla\psi+\nabla\wedge\vec{A}\right)\right]+\vec{G}, \tag{2.42}$$

which describe the propagation of compressional and shear waves, respectively, in an elastic medium.

Under the assumptions made here, the continuity equation (2.28) becomes

$$\dot{\rho}(\vec{x}\,|\,t)+\nabla^2\dot{\psi} = 0. \tag{2.43}$$

Integration of (2.43) with respect to t, gives the perturbation of the density in terms of ψ.

Note that propagation of compressional and shear waves is coupled through the appearance of ψ and $\vec{A}$ in both (2.41) and (2.42), which are scalar and vector wave equations, respectively. If an elastic fluid is being considered then, by definition, no shear stress can be supported. Consequently μ, $\vec{G}$, and hence $\vec{A}$, are all identically zero. The resulting equation for ψ is of the form (2.1) :

$$\lambda\nabla\cdot\left(\rho_0^{-1}\nabla\psi\right)-\ddot{\psi} = -F. \tag{2.44}$$

A model of elastic solids incorporating the effects of a non-zero μ must involve solving both equations simultaneously. Current ability to solve direct problems involving vector wave equations is insufficient to be usefully applied to the associated inverse problems. Boyse and Keller [1986] and Blackledge *et al.* [1987b] employ a slightly more general model, in which non-zero μ is permitted, and gradients of λ and μ are not neglected in (2.36). In both treatments of the associated inverse problem, recourse is made to the Born approximation described in Section 4.2.1, which imposes just as severe a restriction on the validity of their models. In addition, both experimenters eventually assume that the effect of shear waves is negligible or can be identified, so that ultimately only a scalar wave equation is considered.

2.2.3 Acoustic waves

The derivation presented here closely follows Wilcox [1984, Chapter 1], but includes the effect of viscosity [Wall 1987]. The physical parameters characterising a compressible, viscous, heat conducting fluid are pressure $p(\vec{x}\,|\,t)$, density $\rho(\vec{x}\,|\,t)$, velocity $\vec{v}(\vec{x}\,|\,t)$, temperature $\Theta(\vec{x}\,|\,t)$, conductivity $\kappa(\vec{x}\,|\,t)$, and entropy per unit mass $\mathcal{S}(\vec{x}\,|\,t)$. As for elastic waves, motion is described by (2.28) (or, equivalently, (2.29)) and (2.30). In addition, equations describing the thermal properties of the medium are required. These are firstly, an equation relating the variation of $\mathcal{S}$ to the diffusion of heat, and secondly, two equations representing the thermodynamic equation of state. Thus, the field equations are

$$\begin{aligned}
\partial\rho/\partial t + \nabla\cdot(\rho\vec{v}) &= 0, && \text{(2.45a)}\\
\rho D\vec{v}/Dt &= \nabla\cdot\mathcal{T} + \rho\vec{f}, && \text{(2.45b)}\\
\rho D\mathcal{S}/Dt &= \Theta^{-1}\nabla\cdot(\kappa\nabla\Theta), && \text{(2.45c)}\\
\Phi(p,\rho,\Theta) &= 0, && \text{(2.45d)}\\
\phi(p,\rho,\mathcal{S}) &= 0. && \text{(2.45e)}
\end{aligned}$$

The constitutive equation for Newtonian (linear) isotropic viscous fluids relate stress to pressure and shear rate [Spencer 1980, Chapter 8] :

$$\mathcal{T}_{ij} = \mathcal{T}_{ji} = (-p + \lambda\nabla\cdot\vec{v})\,\delta_{ij} + \mu\,(v_{i,j} + v_{j,i})\,, \tag{2.46}$$

where $\lambda(\vec{x})$ and $\mu(\vec{x})$ are the coefficients of viscosity.

Consider, first, a static fluid in which the external body force is zero. Denoting the static fluid parameters by a subscript zero, it is seen that

$$\vec{v}_0 = \vec{f}_0 = 0, \quad \mathcal{T}_0 = -p_0(\vec{x})\mathcal{I}, \tag{2.47}$$

where $\mathcal{I}$ denotes the Kronecker delta. Thus, (2.45b) reduces to

$$\nabla p_0 = 0, \tag{2.48}$$

implying that p_0 is constant. If the static temperature distribution $\Theta_0(\vec{x})$ is specified then $\rho_0(\vec{x})$ and $\mathcal{S}_0(\vec{x})$ are determined by (2.45d) and (2.45e).

Now consider a small perturbation of the static situation. Using ϵ for a small dimensionless parameter, the fluid parameters are written as

$$\begin{aligned} \vec{f}(\vec{x}\,|\,t) &= \epsilon\vec{f}_1(\vec{x}\,|\,t), \\ p(\vec{x}\,|\,t) &= p_0 + \epsilon p_1(\vec{x}\,|\,t), \\ \rho(\vec{x}\,|\,t) &= \rho_0 + \epsilon\rho_1(\vec{x}\,|\,t), \\ \vec{v}(\vec{x}\,|\,t) &= \epsilon\vec{v}_1(\vec{x}\,|\,t), \\ \mathcal{S}(\vec{x}\,|\,t) &= \mathcal{S}_0(\vec{x}) + \epsilon\mathcal{S}_1(\vec{x}\,|\,t). \end{aligned} \tag{2.49}$$

These expressions are now substituted into (2.45a) through (2.45e) and (2.46), terms containing powers of ϵ higher than the first being neglected. In addition, the adiabatic hypothesis is assumed, which stipulates that acoustic disturbances are so rapid that heat diffusion is negligible during the passage of the acoustic waves. Thus, setting $\kappa = 0$ in (2.45c) results in

$$D\mathcal{S}/Dt = 0. \tag{2.50}$$

The local phase velocity may be introduced in the following manner. First, the convective derivative of (2.45e) yields, after invoking (2.50),

$$\frac{\partial\phi}{\partial p}\frac{Dp}{Dt} + \frac{\partial\phi}{\partial\rho}\frac{D\rho}{Dt} = 0. \tag{2.51}$$

By applying the definition (2.27) of convective derivative to p and ρ, the linearised expressions

$$Dp/Dt = \epsilon\partial p_1/\partial t, \tag{2.52}$$

$$D\rho/Dt = \epsilon(\partial\rho_1/\partial t + \vec{v}_1\cdot\nabla\rho_0), \tag{2.53}$$

are obtained. Furthermore, the partial derivatives of ϕ may be expanded in a Taylor series about $(p_0, \rho_0, \mathcal{S}_0)$:

$$\frac{\partial\phi}{\partial p}(p_0 + \epsilon p_1, \rho, \mathcal{S}) = \phi_p^0 + O(\epsilon), \tag{2.54}$$

$$\frac{\partial\phi}{\partial\rho}(p, \rho_0 + \epsilon\rho_1, \mathcal{S}) = \phi_\rho^0 + O(\epsilon), \tag{2.55}$$

where

$$\phi_p^0(\vec{x}) = \frac{\partial\phi}{\partial p}(p_0, \rho_0(\vec{x}), \mathcal{S}_0(\vec{x})), \tag{2.56}$$

and

$$\phi_\rho^0(\vec{x}) = \frac{\partial\phi}{\partial\rho}(p_0, \rho_0(\vec{x}), \mathcal{S}_0(\vec{x})). \tag{2.57}$$

So, to first order in ϵ, (2.51) reduces to

$$\phi_p^0\partial p_1/\partial t + \phi_\rho^0(\partial\rho_1/\partial t + \vec{v}_1\cdot\nabla\rho_0) = 0. \tag{2.58}$$

Defining the local phase velocity $c(\vec{x})$ by

$$\begin{aligned} c^2(\vec{x}) &= -\phi_\rho^0/\phi_p^0 \\ &= \left.\frac{\partial p}{\partial\rho}\right|_{\mathcal{S}_0}, \end{aligned} \tag{2.59}$$

the pressure perturbation is seen to be related to density by

$$\partial p_1/\partial t = c^2\left(\partial \rho_1/\partial t + \vec{v}_1 \cdot \nabla \rho_0\right). \tag{2.60}$$

After similar manipulations to those presented in Section 2.2.2, the continuity equation becomes

$$\begin{aligned} 0 &= \partial \rho_1/\partial t + \nabla \rho_0 \cdot \vec{v}_1 + \rho_0 \nabla \cdot \vec{v}_1 \\ &= c^{-2}\partial p_1/\partial t + \rho_0 \nabla \cdot \vec{v}_1, \end{aligned} \tag{2.61}$$

and the equation of motion becomes

$$\rho_0 \partial \vec{v}_1/\partial t = -\nabla p_1 + \nabla\left[(\lambda + 2\mu)\nabla \cdot \vec{v}_1\right] - \mu \nabla \wedge \nabla \wedge \vec{v}_1 + \rho_0 \vec{f}_1. \tag{2.62}$$

To obtain an equation in terms of velocity, differentiate (2.62) with respect to t and use (2.61) to eliminate p, resulting in

$$\rho_0 \frac{\partial^2 \vec{v}_1}{\partial t^2} = \nabla\left(\rho_0 c^2 \nabla \cdot \vec{v}_1\right) + \frac{\partial}{\partial t}\nabla\left[(\lambda + 2\mu)\nabla \cdot \vec{v}_1\right] - \frac{\partial}{\partial t}\mu \nabla \wedge \nabla \wedge \vec{v}_1 + \frac{\partial}{\partial t}\rho_0 \vec{f}_1. \tag{2.63}$$

In an analogous manner to (2.39), express $\rho_0 \vec{v}_1$ as

$$\rho_0 \vec{v}_1 = \nabla \psi + \nabla \wedge \vec{A}, \tag{2.64}$$

and recall (2.40). Separation of gradient and curl terms in (2.63) yields

$$\frac{\partial^2 \psi}{\partial t^2} = \left[\rho_0 c^2 + (\lambda + 2\mu)\frac{\partial}{\partial t}\right] \nabla \cdot \rho_0^{-1}\left(\nabla \psi + \nabla \wedge \vec{A}\right) + \frac{\partial F}{\partial t}, \tag{2.65}$$

$$\frac{\partial^2 \vec{A}}{\partial t^2} = -\frac{\partial}{\partial t}\mu \nabla \wedge \rho_0^{-1}\left(\nabla \psi + \nabla \wedge \vec{A}\right) + \frac{\partial \vec{G}}{\partial t}. \tag{2.66}$$

Again, in order to neglect vector wave motion, it must be assumed that $\mu = 0$. A simple rearrangement of (2.65) results in an equation of the underlying form (2.1) :

$$(\rho_0 c^2 + \lambda \partial_t)\nabla \cdot (\rho_0^{-1} \nabla \psi) - \partial_{tt}\psi = -\Sigma, \tag{2.67}$$

where

$$\Sigma = \dot{F}. \tag{2.68}$$

The term $\lambda \partial_t$ causes viscous dissipation, which has an effect similar to the diffusion of heat [Morse and Ingard 1968, Section 6.4].

2.3 FREQUENCY-DOMAIN EQUATIONS

Since the underlying time-domain equation is linear in $\psi(x \mid t)$, (but not in the constitutive parameters) a totally equivalent theory can be formulated in terms of the monochromatic temporal frequency components of ψ [Goodman 1968, Chapter 1]. It is convenient to introduce what is here called a preliminary frequency-domain field $\tilde{\psi}(\vec{x}, k)$, which is defined by

$$\tilde{\psi}(\vec{x}, k) = \int_{-\infty}^{\infty} \psi(\vec{x} \mid t) e^{-ikt}\, dt, \tag{2.69}$$

and obeys the inverse relation

$$\psi(\vec{x}\,|\,t) = \int_{-\infty}^{\infty} \tilde{\psi}(\vec{x},k)e^{ikt}\,dk. \tag{2.70}$$

Upon substitution into (2.1), $\tilde{\psi}$ is found to satisfy the underlying frequency-domain equation,

$$\left[\nabla\cdot(\gamma^{-1}\nabla) - \tilde{\mathcal{L}}_0(\vec{x},k)/\tilde{\mathcal{L}}_1(\vec{x},k)\right]\tilde{\psi}(\vec{x},k) = -\tilde{\Sigma}(\vec{x},k)/\tilde{\mathcal{L}}_1(\vec{x},k), \quad \vec{x}\in\Upsilon, \tag{2.71}$$

where

$$\tilde{\mathcal{L}}_1(\vec{x},k) = \sum_{n=0}^{N_1}\alpha_n(\vec{x})(ik)^n, \tag{2.72}$$

$$\tilde{\mathcal{L}}_0(\vec{x},k) = \sum_{n=0}^{N_0}\beta_n(\vec{x})(ik)^n, \tag{2.73}$$

and $\tilde{\Sigma}(\vec{x},k)$ is defined in terms of $\Sigma(\vec{x}\,|\,t)$ in the same way that $\tilde{\psi}(\vec{x},k)$ is defined in terms of $\psi(\vec{x}\,|\,t)$. Recall that (2.5) implies that the free-space wavespeed c is normalised to unity. For other conventions, k is replaced above by the angular frequency $\omega = kc$.

The linearity of (2.1), together with the assumption that the constitutive parameters are time-independent, permits the individual frequency components of the field describing the emanations to be analysed independently. In the time-domain formulation (2.1) the constitutive parameters α_n and β_n, describing the region of interest, appear in differential operators that involve taking time-derivatives of $\psi(\vec{x}\,|\,t)$. By introducing the function $\tilde{\psi}(\vec{x},k)$, which is the (temporal, as opposed to spatial) Fourier transform of $\psi(\vec{x}\,|\,t)$, these parameters now appear in expressions (2.72) and (2.73) which are merely polynomials in k. Thus, all time-derivatives are replaced by a product of the undifferentiated field and the ratio of two polynomials in k. Furthermore, (2.71) contains spatial derivatives of order no greater than two and, since γ is assumed strictly positive (refer to Section 2.2), is an elliptic partial differential equation. Comments on the consequences this imposes for inverse scattering techniques are made in Section 2.6.

2.4 TRANSFORMATIONS TO CANONICAL FORM

Examination of the underlying frequency-domain equation (2.71) reveals that, although the constitutive parameters α_n and β_n appear in a single ratio, the term γ appears separately. It is convenient to perform the following transformation so that all these functions are incorporated in a single term. Make the substitution [Brekhovskikh 1960]

$$\tilde{\psi}(\vec{x},k) = \gamma^{1/2}(\vec{x})\psi(\vec{x},k), \tag{2.74}$$

which implies that

$$\nabla\tilde{\psi} = \gamma^{1/2}\nabla\psi - (1/2)\gamma^{-1/2}\psi\nabla\gamma, \tag{2.75}$$

$$\nabla\cdot(\gamma^{-1}\nabla\tilde{\psi}) = \gamma^{-1/2}\left[\nabla^2 + \frac{\nabla^2\gamma}{2\gamma} - \frac{3\nabla\gamma\cdot\nabla\gamma}{4\gamma^2}\right]\psi. \tag{2.76}$$

Thus, (2.71) can be written as

$$\gamma^{-1/2}\nabla^2\psi + \gamma^{1/2}\left[\frac{\nabla^2\gamma}{2\gamma} - \frac{3\nabla\gamma\cdot\nabla\gamma}{4\gamma^2}\right]\psi - \gamma^{1/2}\frac{\tilde{\mathcal{L}}_0}{\tilde{\mathcal{L}}_1}\psi = -\tilde{\Sigma}/\tilde{\mathcal{L}}_1. \tag{2.77}$$

Upon multiplying through by $\gamma^{1/2}$, the following canonical form is attained :

$$\left[\nabla^2 + k^2\chi^2(\vec{x},k)\right]\psi(\vec{x},k) = -\Sigma(\vec{x},k), \tag{2.78}$$

where

$$\chi^2(\vec{x},k) = \left(\frac{\nabla^2\gamma}{2\gamma} - \frac{3\nabla\gamma\cdot\nabla\gamma}{4\gamma^2} - \frac{\gamma\tilde{\mathcal{L}}_0}{\tilde{\mathcal{L}}_1}\right)/k^2, \tag{2.79}$$

and

$$\Sigma(\vec{x},k) = \gamma^{1/2}\tilde{\Sigma}/\tilde{\mathcal{L}}_1. \tag{2.80}$$

The term χ is seen to be the generalised constitutive parameter referred to in Section 1.4. Substitution of (2.4), (2.72), and (2.73) into (2.79) reveals that the free-space value of χ is unity.

The transformation (2.74) can be applied equally well to $\psi(\vec{x}\,|\,t)$ to produce $\bar{\psi}(\vec{x}\,|\,t)$. When the manipulations corresponding to (2.75) and (2.76) are inserted into (2.1), the following is obtained :

$$\left[\mathcal{L}_1(\vec{x}\,|\,\partial_t)\nabla^2 - \bar{\mathcal{L}}_0(\vec{x}\,|\,\partial_t)\right]\bar{\psi}(\vec{x}\,|\,t) = -\bar{\Sigma}(\vec{x}\,|\,t), \tag{2.81}$$

where

$$\bar{\mathcal{L}}_0(\vec{x}\,|\,\partial_t) = \mathcal{L}_0\gamma + \left(\frac{\nabla^2\gamma}{2\gamma} - \frac{3\nabla\gamma\cdot\nabla\gamma}{4\gamma^2}\right)\mathcal{L}_1, \tag{2.82}$$

and

$$\bar{\Sigma}(\vec{x}\,|\,t) = \gamma^{1/2}\Sigma(\vec{x}\,|\,t). \tag{2.83}$$

It is not possible, however, to divide (2.81) by $\mathcal{L}_1$ which is still a differential operator involving time-derivatives.

From the point of view of this thesis, the time-domain is now sufficiently explored. Because it is argued in Section 2.6 that, although advantages accrue from employment of the time-domain, the frequency-domain leads to greater ease in understanding and solving inverse problems, attention is now concentrated on frequency-domain approaches to direct and inverse scattering.

Inspection of (2.15),(2.44), and (2.67), reveals that χ^2, as given by (2.79), assumes the form

$$\chi^2 = \nu^2 - i\beta/k, \tag{2.84}$$

$$\chi^2 = \frac{\rho_0}{\lambda} + \frac{2\rho_0\nabla^2\rho_0 - 3\nabla\rho_0\cdot\nabla\rho_0}{4\rho_0^2}k^{-2}, \tag{2.85}$$

$$\chi^2 = \frac{\rho_0}{\rho_0 c^2 + ik\lambda} + \frac{2\rho_0\nabla^2\rho_0 - 3\nabla\rho_0\cdot\nabla\rho_0}{4\rho_0^2}k^{-2}, \tag{2.86}$$

for electromagnetic, elastic, and acoustic, media respectively. Only a minority of literature on electromagnetic inverse problems consider the frequency-dependent attenuation term $i\beta/k$. The computational difficulties are so great that often only problems

involving one spatial dimension are tractable [Kristensson and Krueger 1987]. When attenuation is ignored, the canonical equation (2.78) reduces to the inhomogeneous Helmholtz equation

$$\left(\nabla^2 + k^2\nu^2(\vec{x})\right)\psi = 0. \tag{2.87}$$

The constitutive parameter ν describes variations in the refractive index of the medium. The Helmholtz equation also applies to elastic media of constant density, so that $\nu^2 \equiv \lambda$ in (2.85). Perfect acoustic fluids are described by (2.86) with $\lambda = 0$. The most commonly employed model for inverse acoustic scattering assumes that density is constant, leading again to the Helmholtz equation. Although the assumption of constant density is questionable, recent results indicate that variations in density are often of minor importance compared with sound speed variations [Robinson and Greenleaf 1986].

If an elastic medium varies in such a way that $\rho_0(\vec{x}) = \lambda(\vec{x})$ everywhere, then (2.85) reduces to a function of the form [Rose *et al.* 1985]

$$\chi^2(\vec{x}, k) = 1 + V(\vec{x})k^{-2}, \tag{2.88}$$

so that the canonical equation (2.78) becomes

$$\left(\nabla^2 + k^2 + V(\vec{x})\right)\psi = 0. \tag{2.89}$$

The last equation is more familiar as the frequency-domain Schrödinger equation where V represents the scattering potential. Note that the refractive index, which is determined by the part of (2.88) that is independent of k, is unity. Besides being equivalent to the time-dependent Schrödinger equation, the time-domain version of (2.89), obtained by Fourier transformation (see Section 2.3), is also referred to as the plasma-wave equation [Newton 1985], [DeFacio and Rose 1985]. The one-dimensional version of (2.89) has been used to model radio-wave propagation in the ionosphere [Budden 1985,Jordan and Ahn 1979].

Perhaps the two most common forms of underlying equation considered in the inverse scattering literature are the inhomogeneous Helmholtz equation (2.87) and the frequency-domain version (2.89) of the Schrödinger equation. When ν depends on only one spatial variable x, say, it is possible to transform the Helmholtz equation into the Schrödinger equation. Make the change of independent and dependent variables [Olver 1974, Chapter 6], [Bates and Wall 1976] :

$$\hat{x} = \int_{-\infty}^{x} \nu(z)\, dz, \tag{2.90}$$

$$\hat{\psi}(\hat{x}, k) = \nu^{1/2}\psi(x, k), \tag{2.91}$$

and denote $\nu(x)$ in the new coordinate system by $\hat{\nu}(\hat{x})$. Since k is to be treated as an arbitrary, but fixed, constant and because x and $\hat{x}$ are merely scalar variables, gradient and Laplacian operators reduce to ordinary first and second derivatives,

$$\begin{aligned} \psi'(x) &= \nu\frac{d}{d\hat{x}}\left(\nu^{-1/2}\hat{\psi}\right) \\ &= \hat{\nu}^{1/2}\hat{\psi}' - \hat{\nu}^{-1/2}\hat{\nu}'\hat{\psi}/2, \end{aligned} \tag{2.92}$$

and

$$\psi'' = \hat{\nu}^{3/2}\hat{\psi}'' + \frac{(\hat{\nu}')^2 - 2\hat{\nu}\hat{\nu}''}{4\hat{\nu}^{1/2}}\hat{\psi}, \tag{2.93}$$

so that (2.87) becomes

$$\hat{\psi}''(x,k) + \left(k^2 + \hat{V}(x)\right)\hat{\psi}(\hat{x},k) = -\Sigma, \tag{2.94}$$

where

$$V(\hat{x}) = \frac{(\hat{\nu}')^2 - 2\hat{\nu}\hat{\nu}''}{4\hat{\nu}^2} \tag{2.95}$$

and

$$\hat{\Sigma} = \nu^{-3/2}(x)\Sigma. \tag{2.96}$$

The one-dimensional Helmholtz equation has thus been transformed into the Schrödinger equation. Three-dimensional problems involving the Helmholtz equation that are spherically symmetric are, in effect, only one-dimensional since only the radial coordinate is involved [Eftimiu and Huddleston 1984]. Such problems can also be converted to the Schrödinger equation [Prosser 1972]. Provided only one spatial variable is considered, the transformations (2.90) and (2.91) can also be applied to the time-domain versions of (2.87) and (2.89). Thus,

$$\left(\partial_{xx} - \nu^2\partial_{tt}\right)\psi(x\,|\,t) = 0, \tag{2.97}$$

can be converted to

$$\left(\partial_{\hat{x}\hat{x}} - \partial_{tt} + V\right)\hat{\psi}(\hat{x}\,|\,t) = 0. \tag{2.98}$$

In the inhomogeneous Helmholtz equation, waves travel with a velocity $1/\nu(x)$ that varies with position. The transformations (2.90) and (2.91) invoked above define a new spatial coordinate in which waves travel with a constant speed. In one dimension, this is possible since the direction in which the waves are transmitted (and reflected) is known. Bates and Millane [1981] comment on the desirability of generalising the transformation to higher dimensions, but the task proves impossible. The reason for this is that the velocity of a wave passing through a particular point depends on the value of ν at that point, and is also affected by the direction in which the wave is travelling at that instant (as determined by the laws of refraction). It is only within a quasi-particle description of wave propagation [Marcuvitz 1980], [Calderone *et al.* 1985], [Lee and Marcuvitz 1986] or the parabolic approximation [DeSanto 1977], [Corones and Krueger 1983], [Siegmann *et al.* 1985], [Thomson and Wood 1987] to the Helmholtz equation that an equation of the Schrödinger type is obtained. The Helmholtz equation then admits what is called factorisation [Fishman and McCoy 1984a], [Fishman and McCoy 1984b], [Fishman and McCoy 1984c], [Weston 1987].

2.5 SUPPLEMENTARY EQUATIONS

In this thesis, differential equations are understood in the context of generalised functions [Shilov 1968,Stakgold 1979]. In such a setting, classical (i.e. sufficiently continuous and differentiable) solutions are permitted as well as "weak solutions" that exist only in an operational sense. Consequently, precise meaning can be attributed to entities such as the Dirac delta function.

In order to specify a complete description of the behaviour of particular emanations, the differential equations of the previous two sections must be supplemented

by conditions that are entirely dependent on the particular physical situation under consideration. Below are listed typical supplementary conditions. First, boundary conditions are considered. Boundary conditions apply in both time- and frequency-domain formulations. In the former, each is taken to apply for all time. Thus, no explicit functional dependence of ψ is shown. Initial conditions only apply in the time-domain. Radiation conditions are appropriate in unbounded spatial domains.

2.5.1 Boundary conditions

The spatial domain over which the differential equation is defined is denoted D, and its boundary denoted ∂D. Boundary conditions are applicable on ∂D and fall into three classes :

Dirichlet boundary condition :

$$\psi = 0 \text{ on } \partial D. \tag{2.99}$$

Such a condition can be used to describe the vanishing of the tangential electric field on the surface of a perfect conductor. Alternatively, if ψ is the velocity potential for acoustic waves, then (2.99) is appropriate for a completely flexible (non-rigid) boundary, and is often called the "sound soft" boundary condition.

Neumann boundary condition :

$$\partial\psi/\partial n = 0 \text{ on } \partial D, \tag{2.100}$$

where $\partial/\partial n$ denotes the normal derivative on ∂D. This describes the "sound hard" boundary condition, requiring the acoustic velocity across the surface (i.e. in the direction of the normal $\vec{n}$) to vanish identically. It also applies to the tangential magnetic field on the surface of a perfect conductor.

Impedance boundary condition :

$$\partial\psi/\partial n - Z\psi = 0 \text{ on } \partial D. \tag{2.101}$$

Here, Z represents an impedance, and may be a function of position on the boundary. There do not seem to be any physical situations exactly described by an impedance boundary condition, which can nonetheless be convenient and useful, e.g. for highly but not perfectly conducting boundaries [Senior 1960]. When treating inverse problems, the shape of the region D is often unknown. Since Z depends on the region under consideration, it is unlikely that the impedance can be specified *a priori*, and therefore cannot be easily incorporated into an inverse technique. Exceptions to this difficulty are the boundary conditions listed above, which correspond to $Z = \infty$ and $Z = 0$ respectively.

2.5.2 Initial conditions

These specify the field and its derivatives throughout D at some convenient time, $t = 0$, say. In wave problems, it is often more convenient to impose the equivalent condition of causality

$$\psi(\vec{x}\,|\,t) = 0, \quad t < 0. \tag{2.102}$$

2.5.3 Radiation condition

Let S denote a closed surface. The Sommerfeld radiation condition [Müller 1969, Chapter 3],

$$\iint_S |\partial\psi/\partial n + ik\psi| \, dS = o(1) \quad \text{as} \quad S \to \infty, \tag{2.103}$$

ensures that scattered waves travelling in an unbounded region contain a finite amount of energy. Note that the notation $f(x) = o(g(x))$ means [Olver 1974, Chapter 1]

$$\lim_{x\to\infty} f(x)/g(x) = 0, \tag{2.104}$$

which is not to be confused with $f(x) = O(g(x))$ which means [Olver 1974, Chapter 1]

$$\lim_{x\to\infty} f(x)/g(x) = L, \tag{2.105}$$

L being finite in general.

2.5.4 Jump conditions

Classical solutions of a differential equation are required to be sufficiently differentiable. Thus, the classical differential equation only applies in domains where the functions appearing as coefficients are suitably continuous. Weak solutions of a differential equation are not restricted in the same way, but must be constrained across discontinuities by jump conditions. Jump conditions are distinct from boundary conditions because boundary conditions must be imposed independently from the differential equation whereas jump conditions are stipulated by the differential equation.

To derive jump conditions, the differential equation is rewritten in integral form. A differential equation is often derived from an integral expression representing the conservation of some physical quantity (such as mass or charge). The differential equation is merely the manifestation of the integral expression when the volume of integration is of infinitesimal size. Homogeneous jump conditions typically express continuity of some physical quantity across an interface, and take the form

$$[\varrho\psi] = 0, \tag{2.106}$$

$$[\varpi\partial\psi/\partial n] = 0, \tag{2.107}$$

where $[x]$ denotes the difference in value of x on either side of the interface. The forms assumed by ϱ and ϖ, which are simply related to the constitutive parameters of the medium under consideration, for standard mathematical models of physical situations can be found in any number of texts [Jackson 1975, Introduction], [Born and Wolf 1970, Chapter 1], [Jaunzemis 1967, Chapter 3].

Jump conditions cause a significant increase in the complication of integral formulations of scattering, as is shown in Section 3.4, and prohibit simple computational solution of inverse problems [Angell *et al.* 1987,Colton and Monk 1987]. The class of problems governed by the uniform jump conditions,

$$[\varrho] = [\varpi] = 0, \tag{2.108}$$

comfortably avoid the extra complication associated with incorporation of non-uniform jump conditions because both ψ and its normal derivative are continuous everywhere. The conventional volume-source formulation, which encompasses the Born and Rytov approximations (described in Chapter 4) that lead to diffraction tomography (see Section 6.5), implicitly assume that the uniform jump conditions hold.

2.6 COMPARISON OF TIME- AND FREQUENCY- DOMAIN FORMULATIONS

The previous sections in this chapter demonstrate how time- and frequency- domain formulations can be derived to model various physical processes. This section summarizes the advantages and limitations of the respective models.

A feature of formulating scattering problems in the frequency-domain is the ability to pose the problem in terms of a single generalised constitutive parameter χ, as demonstrated in Section 2.4 [Bates 1984]. Phenomena such as dispersion, which require special treatment in time-domain approaches, are easily accounted for by permitting complex and frequency-dependent values of χ. The constitutive parameters appear in χ as coefficients in a polynomial fraction of k. If χ can be determined for all $\vec{x}$ and sufficiently many k, then the simple functional dependence of χ on k permits γ and each α_n, β_n to be extracted. All that is required is to merely solve a set of linear simultaneous equations, obtained at different values of k.

Note, however, that by assuming k to be fixed, the experimenter loses one degree of freedom in the ability to vary the incident field. Inverse scattering algorithms developed in the frequency-domain usually gather data from remote measurements performed at a variety of combinations of transmitter/receiver positions. In time-domain approaches, where the frequency of the incident field can be varied, it is often sufficient to keep either the transmitter or the receiver position fixed.

The major conceptual distinction between time- and frequency- domain formulations can be understood by noting that the underlying frequency-domain equation is an elliptic partial differential equation whereas the time-domain equation is hyperbolic [Garabedian 1964, Chapter 3]. Partial differential equation theory reveals that solutions to elliptic equations are "better-behaved" than their hyperbolic counterparts. Whereas solutions to elliptic equations are "smooth", hyperbolic equations admit the propagation of discontinuities [John 1978, Chapter 5]. This property of hyperbolic equations is most often exploited in time-domain approaches by employing the notion of a pulse of radiation that exists for a finite duration. Hyperbolic differential equations also admit the concept of a domain of dependence [Garabedian 1964, Chapter 5]. The ability of a disturbance at $(\vec{x}\,|\,t)$ to remain unnoticed at $(\vec{x}_1\,|\,t_1)$ until $|t_1 - t| > |\vec{x}_1 - \vec{x}|$ forms the essence of the concept of causality. This concept is sacrificed in the frequency domain since monochromatic waves with the implicit temporal dependence $\exp(ikt)$ necessarily exist for all time, thereby excluding the notion of a pulse. The notion of causality is a powerful argument for the preferability of time-domain approaches, but is also a major weakness, due to the numerical instabilities inherent in solving problems involving discontinuities.

Although from an analytical perspective, time- and frequency- domain approaches are equivalent, and despite the conceptual advantages of invoking the concept of causality, the numerical difficulties arising when solving scattering problems in the time-domain are daunting. The above arguments suggest that it is more profitable to perform analysis in the frequency-domain. Consequently, the remaining chapters develop solutions to both direct and inverse problems using (2.78) as a starting point.

The above conclusion does not hold for scattering in a single space dimension or, equivalently, for scatterers whose constitutive parameters vary in only one

dimension. This is because, as demonstrated in Section 2.4, one can readily transform between the one-dimensional Helmholtz and Schrödinger equations. From the point of view of this thesis, however, the above remark does not invalidate the conclusion drawn in the previous paragraph, since the aim (as stated at the beginning of the Preface) is to provide a general overview of inverse scattering, and one-dimensional scatterers are highly specialised, arising only rarely (even approximately) in the real-world. Consequently, one-dimensional scatterers are only discussed briefly in this thesis, in Sections 2.4 and 6.8.

2.7 THE FOURIER TRANSFORM

The description of sinusoidally varying time waveforms by employing the frequency-domain is universally understood for all linear, time-invariant, systems (refer to Section 2.3). An analogous decomposition of a spatially varying function into spatial frequency components, each with a definite spatial periodicity, introduces the concept of multi-dimensional Fourier transforms. There are many notations for the Fourier transform of a function [Bracewell 1978, Chapter 2]. In this thesis, a function of $\vec{x}$ is generally written as a lower-case Roman or Greek letter, for example $f(\vec{x})$ or $\psi(\vec{x}, k)$, while its Fourier transform is denoted either by $\mathcal{F}[f(\vec{x})](\vec{u})$ or by the corresponding upper-case letter, e.g. $F(\vec{u})$ or $\Psi(\vec{u}, k)$, with the arguments omitted whenever the meaning is clear. The Fourier transform variable $\vec{u} = (u_1, \ldots, u_K)$ corresponding to $\vec{x}$ is, in two spatial dimensions, more conventionally denoted (u, v) (refer to Section 2.2). Image-function space consists of all functions $f(\vec{x})$ and their Fourier transforms $F(\vec{u})$ constitute Fourier-function space. All integrals in this section are understood to extend from $-\infty$ to ∞ for each component of $\vec{x}$ or $\vec{u}$.

The Fourier transform of f is defined as [Bracewell 1978, Chapter 2]

$$\begin{aligned} F(\vec{u}) &= \mathcal{F}[f(\vec{x})](\vec{u}) \\ &= \int f(\vec{x}) e^{-i2\pi\vec{u}\cdot\vec{x}}\, d\vec{x}. \end{aligned} \tag{2.109}$$

The inverse relation is given by

$$\begin{aligned} f(\vec{x}) &= \mathcal{F}^{-1}[f(\vec{u})](\vec{x}) \\ &= \int F(\vec{u}) e^{i2\pi\vec{u}\cdot\vec{x}}\, d\vec{u} \\ &= \mathcal{F}\left[\mathcal{F}[f](\vec{u})\right](-\vec{x}). \end{aligned} \tag{2.110}$$

The classical condition for existence of the Fourier transform is that $|f|^2$ is Lebesgue integrable, although this can be extended in the theory of distributions (see, for example, Arsac [1966] or Shilov [1968, Chapter 2]).

The Fourier transform of a real function is conjugate-symmetric [Bracewell 1978, Chapter 2],

$$[\mathcal{F}[f](-\vec{u})]^* = \mathcal{F}[f](\vec{u}). \tag{2.111}$$

The effect of translating a function is to introduce a linear phase shift to its Fourier transform

$$\mathcal{F}[f(\vec{x} - \vec{a})](\vec{u}) = e^{-i2\pi\vec{u}\cdot\vec{a}} \mathcal{F}[f(\vec{x})](\vec{u}). \tag{2.112}$$

The convolution of two functions, $g(\vec{x})$ and $h(\vec{x})$, is defined by [Bracewell 1978, Chapter 3],

$$\begin{aligned} f(\vec{x}) &= g(\vec{x}) \odot h(\vec{x}) \\ &= \int g(\vec{y}) h(\vec{x} - \vec{y})\, d\vec{y} \\ &= h(\vec{x}) \odot g(\vec{x}), \end{aligned} \tag{2.113}$$

where the last line follows by making the change of variables $\vec{z} = \vec{x} - \vec{y}$. The Fourier transform of a convolution is [Bracewell 1978, Chapter 6],

$$\mathcal{F}[f] = \mathcal{F}[g]\mathcal{F}[h]. \tag{2.114}$$

The Dirac δ-function $\delta(\vec{x})$ is defined in a distributional sense by

$$\int f(\vec{x})\delta(\vec{x})\, d\vec{x} = f(\vec{0}), \tag{2.115}$$

where $\vec{0}$ is the zero vector. Clearly, from the properties of the δ-function,

$$f(\vec{x}) = f(\vec{x}) \odot \delta(\vec{x}), \tag{2.116}$$

holds for any $f(\vec{x})$.

The autocorrelation of f is defined as

$$ff(\vec{x}) = f(\vec{x}) \odot f^*(-\vec{x}). \tag{2.117}$$

When $\vec{u}$ is permitted to have complex-valued components, ff has the Fourier transform

$$\begin{aligned} \mathcal{F}[ff(\vec{x})] &= \iint f^*(\vec{y}) f(\vec{x} + \vec{y}) e^{-i2\pi\vec{u}\cdot\vec{x}}\, d\vec{x}\, d\vec{y} \\ &= \iint f^*(\vec{y}) f(\vec{z}) e^{-i2\pi\vec{u}\cdot(\vec{z}-\vec{y})}\, d\vec{z}\, d\vec{y} \\ &= \int f^*(\vec{y}) e^{i2\pi\vec{u}\cdot\vec{y}}\, d\vec{y} \int f(\vec{z}) e^{-i2\pi\vec{u}\cdot\vec{z}}\, d\vec{z} \\ &= \left[\mathcal{F}[f](\vec{u^*})\right]^* \mathcal{F}[f](\vec{u}). \end{aligned} \tag{2.118}$$

In the event that the components of $\vec{u}$ are real-valued, $\vec{u^*} = \vec{u}$, the Fourier transform of ff is entirely real, i.e.

$$\mathcal{F}[ff](\vec{u}) = |\mathcal{F}[f](\vec{u})|^2. \tag{2.119}$$

Denote by $supp(f)$, the support of f, i.e. the set of points at which f is non-zero,

$$supp(f) = \{\vec{x} : f(\vec{x}) \neq 0\}. \tag{2.120}$$

A function f is said to be of finite extent (compact support) if the linear extent of $supp(f)$ in each Cartesian direction, L_i say, is finite,

$$f(\vec{x}) = 0 \quad \text{if any } |x_i| > L_i/2. \tag{2.121}$$

For functions that satisfy positivity, i.e.

$$f(\vec{x}) \geq 0 \quad \text{for all } \vec{x}, \tag{2.122}$$

ff is also positive and has a linear extent in each Cartesian direction, LL_i, double that of f,

$$LL_i = 2L_i. \tag{2.123}$$

The position has now been reached where it is appropriate to provide a concrete example of the manner in which a function can be represented by an expansion in terms of basis functions (refer to Section 1.5). Consider a two-dimensional function $f(x,y)$ of finite extent. The discrete Fourier series representation for f within $supp(f)$ is

$$f(\vec{x}) = \frac{1}{L_1 L_2} \sum_{m_1=-M_1}^{M_1} \sum_{m_2=-M_2}^{M_2} F_{m_1 m_2} e^{-i2\pi(\frac{m_1 x}{L_1} + \frac{m_2 y}{L_2})}. \tag{2.124}$$

The exponential functions are to be associated with the Ξ_m of Section 1.5 (after suitable re-ordering of the pair of integers m_1, m_2) while the quantities $F_{m_1 m_2}$ are the expansion coefficients. The sampling theorem [Bates and McDonnell 1986, Section 10] states that the $F_{m_1 m_2}$ are merely samples of the Fourier transform of f spaced at intervals of $1/L_1$ and $1/L_2$ in the u and v directions respectively, i.e.

$$f_{m_1 m_2} = F(m_1/L_1, m_2/L_2). \tag{2.125}$$

For general f, M_1 and M_2 are infinite. Truncation to finite values of M_1 and M_2 is not only necessary for the reasons intimated in Section 1.5, but also represents a procedure for overcoming ill-posedness, for the $|F_{m_1 m_2}|$ associated with physically meaningful functions f inevitably decrease with increasing m_1 and m_2 to a level comparable to the noise present in the system. The intricacies of dealing with infinite summations becomes apparent when it is realised that, with the 2-norm (the energy norm) and the ∞-norm (the maximum norm) defined respectively by

$$\|f\|_2 = \left[\int |f(\vec{x})|^2 \, d\vec{x}\right]^{1/2}, \tag{2.126}$$

$$\|f\|_\infty = \sup_{\vec{x}} |f(\vec{x})|, \tag{2.127}$$

(2.124) converges to f in the former but not the latter. This is commonly known as Gibb's phenomenon [Bracewell 1978, Chapter 10].

Chapter 3

WAVE PROPAGATION CONCEPTS AND TECHNIQUES

3.1 INTRODUCTION

The canonical frequency-domain equation (2.78), together with appropriate supplementary equations (see Section 2.5), describes the underlying relationship between the generalised constitutive parameter χ and the field ψ characterising the emanations. Explicit analytical expressions for ψ in terms of readily evaluable functions exist only for a limited class of geometries [Bowman *et al.* 1969]. Typically, the boundary of χ must coincide with a coordinate surface in one of the "separable" coordinate systems [Arfken 1970, Chapter 2], and variation in χ is restricted to a very small number of parameters, e.g. a constant value within the radius of a single cylinder.

For a general χ, it is difficult to establish explicit expressions for ψ. From the differential equation, it is not apparent how changes in χ affect ψ. Physical insight concerning the radiation process can be gained by reformulating the differential equation as an integral equation, in which the inhomogeneity is replaced by equivalent sources that radiate according to some propagation principle. Before presenting general integral formulations in Section 3.4, it is appropriate to introduce some preliminary notions.

In Section 3.2 are presented some common propagation principles that have been successfully applied to direct problems, and may lead to useful inverse techniques.

A second fundamental concept is that of a Green's function. The development of Green's functions presented in Section 3.3 has slightly different emphasis to that usually appearing in scattering theory.

In Section 3.4, Green's (second) theorem is invoked to obtain a general integral relation between the generalised constitutive parameter and the field, from which are derived several expressions that are of use in direct and inverse scattering problems. One of these is further developed in Appendix 3-A, which is a generalisation [Tan and Bates 1988] of the volume-source [Jones 1964, Section 6.13], or polarisation-source [Bates and Ng 1972], principle.

Integral formulations of inverse scattering tend to be adopted in favour of differential formulations, for which there appears to be two plausible reasons. First, as is shown by the development presented in this chapter, they permit incorporation of whatever propagation principle is considered appropriate for a particular situation. A

second advantage of integral over differential formulations is that ψ and χ appear explicitly in integrals, which are more readily susceptible to analytical manipulation and approximations. The integral relations of Section 3.4 foreshadow a detailed discussion in Chapter 4 of various approaches to the direct problem.

3.2 TRAVELLING WAVES AND THEIR GENERALISATIONS

Consider, first, the canonical frequency-domain equation (2.78), without the source term Σ and with the generalised constitutive parameter taken to be independent of frequency (and therefore written more conventionally as $\nu(\vec{x})$), i.e.

$$(\nabla^2 + k^2\nu^2)\psi(\vec{x}, k) = 0, \quad \vec{x} \in \Upsilon. \tag{3.1}$$

The equation governing wave propagation in free-space, where the generalised constitutive parameter is everywhere unity (refer to Section 2.4), is

$$(\nabla^2 + k^2)\psi(\vec{x}, k) = 0. \tag{3.2}$$

For any vector $\vec{k}$ that obeys $\vec{k} \cdot \vec{k} = k^2$, the function

$$\psi(\vec{x}, k) = e^{-i\vec{k}\cdot\vec{x}}, \tag{3.3}$$

is a solution of (3.2) throughout the whole of Υ. The time-domain equivalent of (3.3) (obtained by Fourier transform — see Section 2.3),

$$\psi(\vec{x} \,|\, k) = \delta(t - \hat{k} \cdot \vec{x}), \tag{3.4}$$

represents an impulsive plane wave travelling in the direction of the unit vector, $\hat{k} = k^{-1}\vec{k}$. For (3.3), surfaces of constant phase,

$$\vec{k} \cdot \vec{x} = \text{ const}, \tag{3.5}$$

coincide with surfaces of constant amplitude.

When the refractive index is inhomogeneous, so that ν varies throughout Υ_s, the wave-front of a forward travelling wave is retarded by different amounts depending on the variation of the inhomogeneity. Consequently, it is preferable to express the field as [Jones 1964, Chapter 6]

$$\psi(\vec{x}, k) = A(\vec{x}, k)e^{-ikS(\vec{x},k)}, \tag{3.6}$$

where wave-fronts are defined to be surfaces on which S is constant. The functions A and S are determined by substitution of (3.6) into (3.1) and collecting like powers of k :

$$\begin{aligned} 0 \;=\; & k^2(\nu^2 - \nabla S \cdot \nabla S)A \\ & - ik(A\nabla^2 S + 2\nabla A \cdot \nabla S) \\ & + \nabla^2 A. \end{aligned} \tag{3.7}$$

If k is large, the most significant term in (3.7) is the coefficient of k^2. Equating this coefficient to zero results in the eikonal equation [Born and Wolf 1970, Chapter 3] :

$$\nabla S \cdot \nabla S = \nu^2. \tag{3.8}$$

Along with initial conditions, (3.8) determines S in terms of ν.

A geometrical ray $\vec{r}(l)$ is defined to be any curve everywhere perpendicular to the geometrical wave-fronts, and represents the direction in which energy propagates. The problem of computing a ray path can be formulated in several ways. One common method invokes arc length as the parameter along a particular ray. The differential equation for a ray becomes [Born and Wolf 1970, Chapter 3]

$$\frac{d}{dl}\left(\nu \frac{d\vec{r}}{dl}\right) = \nabla \nu. \tag{3.9}$$

Ray tracing algorithms [Fawcett and Keller 1985,Lawrence 1985] determine a particular ray by solving (3.9) together with equations specifying the initial direction of the ray.

An alternative method for computing a ray can be formulated after introduction of the point characteristic function [Born and Wolf 1970, Chapter 4],

$$\begin{aligned} V(\vec{x}_1, \vec{x}_0) &= \int_{l_0}^{l_1} \nu\left(\vec{r}(l)\right) dl \\ &= S(\vec{x}_1) - S(\vec{x}_0), \end{aligned} \tag{3.10}$$

which defines the optical length of the ray $\vec{r}(l)$ joining the points $\vec{x}_0 = \vec{r}(l_0)$ and $\vec{x}_1 = \vec{r}(l_1)$. Fermat's principle of least action, together with (3.10), is often used as the foundation of geometrical optics : a ray is defined as a curve for which RHS (3.10) is extremal, with respect to other paths joining the same two points [Cornbleet 1983]. The rays, as defined by (3.9), are also solutions to the Euler–Lagrange equations [Fox 1950] for the variational problem.

It is convenient to introduce Cartesian coordinates (ξ, η), whose origin O coincides with that of the fixed (x, y) Cartesian system, that are rotated with respect to O so that the η-axis runs parallel to the straight line joining the points $\vec{x}_0$ and $\vec{x}_1$, which fall on the line $\xi = \xi_0$. Denote by ϕ the angle formed by the positive ξ- and x- axes (refer to Figure 3.1). When the refractive index variation is moderate enough that the ray passing through $\vec{x}_0$ and $\vec{x}_1$ is directed predominantly in the η-direction, it is appropriate to parametrise the ray by

$$\xi = \xi(\eta), \tag{3.11}$$

$$\xi(\eta_0) = \xi(\eta_1) = \xi_0, \tag{3.12}$$

where, in the rotated coordinate system, $\vec{x}_0 = (\xi_0, \eta_0)$ and $\vec{x}_1 = (\xi_0, \eta_1)$. Under this parametrisation, and using subscripts to denote partial differentiation but $'$ to denote total differentiation, the ray equation becomes [Norton and Linzer 1982]

$$\nu \xi'' = (1 + \xi'^2)(\nu_\eta - \xi' \nu_\xi), \tag{3.13}$$

which must be solved together with (3.12).

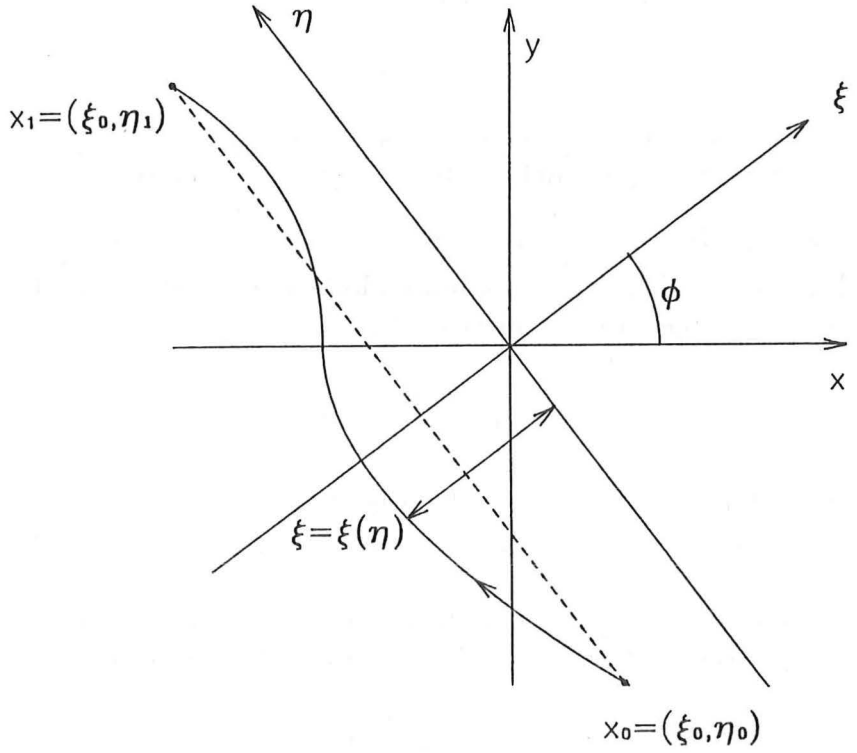

Figure 3.1: Rotated Cartesian coordinates

Returning to (3.7), the term in k yields the first transport equation,

$$A\nabla^2 S + 2\nabla A \cdot \nabla S = 0. \tag{3.14}$$

In view of the eikonal equation (3.8),

$$\nabla S = \nu \hat{l}, \tag{3.15}$$

where $\hat{l}$ denotes the unit tangent vector at $\vec{x}$ of the ray emanating from $\vec{x}_0$. Using $\partial/\partial l$ to denote differentiation along the ray, (3.14) becomes

$$A\nabla\cdot(\nu\hat{l}) + 2\partial A/\partial l = 0. \tag{3.16}$$

When only one spatial dimension x, say, is involved, the tangent vector to the ray is always in the x-direction, and (3.16) reduces to

$$A\nu' + 2A'\nu = 0, \tag{3.17}$$

where $'$ denotes total differentiation with respect to x. Integration of (3.17) yields [Jones 1964, Chapter 6]

$$A = C\nu^{-1/2}, \tag{3.18}$$

C being an arbitrary constant.

For a general inhomogeneous medium that depends on more than one spatial variable, it is difficult to simplify the first term of (3.16). The expression (3.18) is only an approximate solution to (3.16) whose error is tolerable, by some subjective criterion,

for a limited class of refractive index variations. Neither, in general, does the third term in (3.7) vanish identically. Consequently, an expression of the form (3.6) is not a true solution of (3.1). A more accurate solution can be obtained by expressing $\psi(\vec{x}, k)$ as an asymptotic expansion in inverse powers of k [Born and Wolf 1970, Chapter 3],

$$\psi(\vec{x}, k) = e^{-ikS(\vec{x})} \sum_{n=0}^{\infty} A_n(\vec{x})(ik)^{-n}, \tag{3.19}$$

which is particularly appropriate at high frequencies (i.e. when k is large), and provides a better description of the field in the forward direction.

The generalisation of the ray principle to an inhomogeneous medium, in which χ is frequency-dependent and complex-valued, is encompassed in evanescent wave theory, or complex ray analysis [Ghione *et al.* 1984]. The point characteristic function is redefined as

$$V(\vec{x}_1, \vec{x}_0, k) = \int_{l_0}^{l_1} \chi(\vec{r}(l), k)\, dl. \tag{3.20}$$

Provided the real part of χ is large compared with the imaginary part, the propagation is predominantly wave-like. The major effect of the imaginary part is to cause attenuation. In particular, the ray paths are still specified by (3.9) [Einziger and Felsen 1982]. In an analogous manner to (3.8) [Born and Wolf 1970, Chapter 4],

$$\nabla V \cdot \nabla V = \chi^2(\vec{x}, k). \tag{3.21}$$

Although it is possible to develop expressions analogous to (3.19) that better approximate the field in the forward direction, it is often more important to introduce terms to explicitly account for reflection and diffraction. A procedure for doing this is developed in the Section 3.4.1.

3.3 GREEN'S FUNCTIONS

A function $g(\vec{x}, \vec{x}_1, k)$ that is a solution to the equation

$$(\nabla^2 + \alpha(\vec{x}, k) + \beta(\vec{x}, \vec{x}_1, k))g(\vec{x}, \vec{x}_1, k) = -\frac{\delta(\vec{x} - \vec{x}_1)}{A(\vec{x}_1, k)}, \tag{3.22}$$

is said to be a Green's function for the operator $(\nabla^2 + \alpha + \beta)$. The quantities α and β are general functions which are specialised later in this section for particular Green's functions. The amplitude factor $A(\vec{x}_1, k)$ is necessary to account for the dependence of the Green's function on the variation of the medium, as shown by (3.44) below. Throughout this section, $\vec{x}_1$ is arbitrary, but fixed, and all derivatives are with respect to $\vec{x}$. For the situations envisaged in later chapters, Υ is an unbounded domain, implying the need to impose the radiation condition (2.103), so that g represents an outgoing field emanating from a point source at $\vec{x}_1$.

Two forms of Green's function are now introduced.

3.3.1 Conventional Green's function

The most familiar form for g is obtained by setting

$$\alpha = k^2, \tag{3.23}$$

$$\beta = 0, \tag{3.24}$$
$$A = 1. \tag{3.25}$$

The Green's function then satisfies the Helmholtz equation with constant wavenumber and is the familiar free-space Green's function [Jones 1964, Section 1.34]

$$g = \begin{cases} e^{-ikr}/(4\pi r), & \vec{x} \in \mathbf{R}^3, \\ (-i/4)H_0^{(2)}(kr), & \vec{x} \in \mathbf{R}^2, \end{cases} \tag{3.26}$$

where $r = |\vec{x} - \vec{x}_1|$. The far-field form of g occurs when $R = |\vec{x}|$ is large enough that g is adequately represented by [Abramowitz and Stegun 1965, Chapter 9]

$$g \sim \begin{cases} e^{-ik(R-\vec{x}_1\cdot\hat{x})}/(4\pi R), & \vec{x} \in \mathbf{R}^3, \\ e^{-ik(R-\vec{x}_1\cdot\hat{x})}/(8\pi kR)^{1/2}, & \vec{x} \in \mathbf{R}^2, \end{cases} \tag{3.27}$$

where $\hat{x} = \vec{x}/R$. Apart from the amplitude attenuation factor (depending only on R and k) needed to maintain conservation of energy, g behaves far away from the source as a plane wave, $\exp(-ik(R - \vec{x}_1 \cdot \hat{x}))$. In any particular direction, the wave-fronts are given approximately by (3.5). This is because, at great distances, the curvature of spherical wave-fronts becomes negligible.

The conventional Green's function describes radiation from a point source in an isotropic uniform medium, for which there is no reason to expect non-symmetric behaviour. Indeed, g depends only on the distance $|\vec{x} - \vec{x}_1|$ between two points, and not on the actual points themselves. Note, also, that $g(\vec{x}, \vec{x}_1, k) = g(\vec{x}_1, \vec{x}, k)$.

The 11 separable coordinate systems [Morse and Feshbach 1953, Chapter 5] deserve special mention. A general right-handed separable coordinate system is denoted by (ξ_1, ξ_2, ξ_3) where ξ_1 is the radial coordinate and ξ_2 and ξ_3 are angular coordinates. In such coordinates, it is possible to form a set of functions [Morse and Feshbach 1953, page 891–893]

$$Y_m(\xi_2, \xi_3) = X_{2n}(\xi_2)X_{3p}(\xi_3), \tag{3.28}$$

that are mutually orthonormal with respect to some weight function $w(\xi_2, \xi_3)$,

$$\iint Y_m Y_n^* w \, d\xi_2 \, d\xi_3 = \delta_{mn}, \tag{3.29}$$

the integral extending over a surface of constant ξ_1. The indices n and p in (3.28) are related to the separation constants, but it is not necessary to explicitly include the functional dependence of (3.28) on the wavenumber k, which is assumed constant, or on the separation constants. The relevant point is that n and p, and hence m, can be associated with the integers. The set $\{Y_m\}$ also forms a complete set, or basis, for functions of ξ_2 and ξ_3. An arbitrary function $f(\xi_2, \xi_3)$ can be expressed as

$$f(\xi_2, \xi_3) = \sum_m f_m Y_m(\xi_2, \xi_3), \tag{3.30}$$

where the basis, or expansion, coefficients are given by

$$f_m = \iint f(\xi_2, \xi_3) Y_m^*(\xi_2, \xi_3) w(\xi_2, \xi_3) \, d\xi_2 \, d\xi_3, \tag{3.31}$$

on account of (3.29).

Associated with each Y_m are solutions, $y_m^{(r)}(\xi_1)Y_m$ and $y_m^{(o)}(\xi_1)Y_m$, of the Helmholtz equation that are, respectively, standing waves regular (finite) at the origin and outward travelling waves at infinity. (Standing waves infinite at the origin and inward travelling waves can also be derived.) The free-space Green's function can be expressed as [Morse and Feshbach 1953, Section 7.2]

$$g(\xi_1,\xi_2,\xi_3,\xi_1',\xi_2',\xi_3',k) = \sum_m c_m(\xi_2',\xi_3')w(\xi_2',\xi_3')Y_m^*(\xi_2',\xi_3')Y_m(\xi_2,\xi_3)y_m^{(r)}(\xi_{1<})y_m^{(o)}(\xi_{1>}) \tag{3.32}$$

where $\xi_{1<}$ and $\xi_{1>}$ are, respectively, the lesser and greater of ξ_1 and ξ_1'. Note that each c_m only depends on ξ_2' and ξ_3'. The forms for $\vec{\xi}$ and $\{h_i\}$ are tabulated in Chapter 5 of Morse and Feshbach [1953], wherein the weight function w, $\{N_m\}$, and $\{Y_m\}$ are discussed in general terms in Section 7.2, and those that arise in particular separable coordinate systems are dealt with in Chapter 11. For reference in future chapters, the expressions appropriate in cylindrical polar coordinates $(\rho;\phi,z)$ for fields independent of z are [Morse and Feshbach 1953, Section 7.2]

$$Y_m(\phi) = e^{im\phi}, \tag{3.33}$$

$$w(\phi) = 1/2\pi, \tag{3.34}$$

$$g(\rho;\phi,\rho';\phi') = (-i/4)\sum_{m=-\infty}^{\infty} J_m(k\rho_<)H_m^{(2)}(k\rho_>)Y_m(\phi-\phi'). \tag{3.35}$$

Note that the simple addition theorems

$$Y_m(\phi)Y_n(\phi) = Y_{m+n}(\phi), \tag{3.36}$$

$$Y_m(\phi+\phi') = Y_m(\phi)Y_m(\phi'), \tag{3.37}$$

do not hold in more general coordinates.

3.3.2 LJWKB Green's function

When χ is inhomogeneous, it might be expected that a more physically representative description of propagation must incorporate refraction. The eikonal (3.20) discussed in Section 3.2 provides such a description. The Liouville-JWKB Green's function is now defined, in three (spatial) dimensions, as

$$g^{(WKB)}(\vec{x},\vec{x}_1,k) = \frac{\chi^{-1/2}(\vec{x},k)e^{-ikV(\vec{x},\vec{x}_1,k)}}{4\pi V(\vec{x},\vec{x}_1,k)}. \tag{3.38}$$

Elementary algebra reveals that, when $\vec{x} \neq \vec{x}_1$,

$$\nabla g^{(WKB)} = -\left(\nabla \ln \chi^{1/2} + ik\nabla V + \nabla \ln V\right) g^{(WKB)}, \tag{3.39}$$

$$\nabla^2 g^{(WKB)} = -(\alpha(\vec{x},k) + \beta(\vec{x},\vec{x}_1,k))g^{(WKB)}, \tag{3.40}$$

in accordance with (3.22), when $\alpha(\vec{x},k)$ is chosen to be those terms in $\nabla^2 g^{(WKB)}$ that are independent of $\vec{x}_1$ and $\beta(\vec{x},\vec{x}_1,k)$ is chosen to be the remaining terms :

$$\alpha^{(WKB)} = k^2\chi^2 - \nabla \ln \chi^{1/2} \cdot \nabla \ln \chi^{1/2} + \nabla^2 \ln \chi^{1/2}, \tag{3.41}$$

$$\beta^{(WKB)} = (1+ikV)(\nabla^2 \ln V - \nabla \ln V \cdot \nabla \ln V - \nabla \ln \chi \cdot \nabla \ln V). \tag{3.42}$$

Provided χ is smoothly varying, the Liouville-JWKB Green's function near $\vec{x} = \vec{x}_1$ behaves as

$$g^{(WKB)}(\vec{x}, \vec{x}_1, k) \sim \frac{\chi^{-1/2}(\vec{x}_1, k) e^{-ikr\chi(\vec{x}_1, k)}}{4\pi r \chi(\vec{x}_1, k)}. \tag{3.43}$$

Comparison of (3.43) with (3.26) reveals that

$$A^{(WKB)}(\vec{x}, k) = \chi^{3/2}(\vec{x}, k). \tag{3.44}$$

3.4 INTEGRAL FORMULATIONS

With appropriate restrictions on the function f, and on the domain D whose boundary is denoted S, Green's second theorem is [John 1978, Chapter 3]

$$\int_D f \nabla^2 g \, dV = \int_D g \nabla^2 f \, dV + \int_S \left(f \frac{\partial g}{\partial n} - g \frac{\partial f}{\partial n} \right) dS. \tag{3.45}$$

Here, g is taken to be an arbitrary Green's function, obeying (3.22). To obtain an expression relating the field and the generalised constitutive parameter that incorporates the propagation principle governing g, the canonical frequency-domain equation (2.78) is rewritten as

$$(\nabla^2 + \alpha)\psi = -\Sigma + (\alpha - k^2\chi^2)\psi. \tag{3.46}$$

If $\vec{x}_1$ is in the volume D then substitution of (3.22) and (3.46) into (3.45) yields

$$\begin{aligned} \psi(\vec{x}_1)/A(\vec{x}_1, k) \;=\; & \int_D \Sigma(\vec{x}, k) g(\vec{x}, \vec{x}_1, k) \, dV(\vec{x}) \\ & + \int_D [k^2\chi^2(\vec{x}, k) - \alpha(\vec{x}, k) - \beta(\vec{x}, \vec{x}_1, k)] g(\vec{x}, \vec{x}_1, k) \psi(\vec{x}, k) \, dV(\vec{x}) \\ & + \int_S \left(\psi(\vec{x}, k) \frac{\partial g}{\partial n}(\vec{x}, \vec{x}_1, k) - g(\vec{x}, \vec{x}_1, k) \frac{\partial \psi}{\partial n}(\vec{x}, k) \right) dS(\vec{x}). \end{aligned} \tag{3.47}$$

For inverse scattering problems, the term Σ represents the source of radiation completely under the control of the experimenter. It is appropriate to consider this source as generating the incident field $\psi_i(\vec{x}, k)$ that is used to probe the object of interest,

$$\psi_i(\vec{x}, k) = \int_D A(\vec{x}_1, k) \Sigma(\vec{x}_1, k) g(\vec{x}_1, \vec{x}, k) \, dV(\vec{x}_1), \tag{3.48}$$

and to decompose the total field into the sum of the incident field and a scattered field $\psi_s(\vec{x}, k)$,

$$\psi(\vec{x}, k) = \psi_i(\vec{x}, k) + \psi_s(\vec{x}, k). \tag{3.49}$$

There is no approximation involved in (3.49), which can be viewed as defining ψ_s in terms of ψ and ψ_i. By appropriate selection of the source term Σ, any desired incident field can be used to probe the inhomogeneity [Moses and Prosser 1986]. When (3.47) is multiplied by $A(\vec{x}_1, k)$, and the roles of integration variable and independent variable are interchanged, the definition of the incident field (3.48) results in [Tan and Bates 1988]

$$\begin{aligned}\psi(\vec{x},k) &= \psi_i(\vec{x},k) \\ &\quad + A(\vec{x},k)\int_D [k^2\chi^2(\vec{x}_1,k) - \alpha(\vec{x}_1,k) - \beta(\vec{x}_1,\vec{x},k)]g(\vec{x}_1,\vec{x},k)\psi(\vec{x}_1,k)\,dV(\vec{x}_1) \\ &\quad + A(\vec{x},k)\int_S \left(\psi(\vec{x}_1,k)\frac{\partial g}{\partial n}(\vec{x}_1,\vec{x},k) - g(\vec{x}_1,\vec{x},k)\frac{\partial \psi}{\partial n}(\vec{x}_1,k)\right) dS(\vec{x}_1). \quad (3.50)\end{aligned}$$

Examination of (3.49) reveals that the second and third terms of RHS (3.50) comprise the scattered field.

It is worth noting at this point that the corresponding time-domain manipulations are somewhat less general because it is necessary to collect the highest order time-derivatives in a term $\alpha(\vec{x} \mid \partial_t)$, thereby restricting the freedom to choose a particular form for g. An exception is the plasma wave equation (2.98) in more than one dimension, for which the free-space Green's function can be employed. The equivalent of (3.50) then becomes the Lippmann-Schwinger equation [Newton 1966, Chapter 10].

Three forms of (3.50) are now considered.

3.4.1 Generalised volume-source formulation

Suppose the inhomogeneity is continuous not only throughout Υ_s, but throughout the whole of Υ. That is, χ must be unity on the boundary of Υ_s. It follows that the field and its first derivative are continuous throughout Υ, i.e. ψ satisfies the uniform jump conditions (see Section 2.5.4). The domain D in (3.50) can be taken to be the whole of Υ. The boundary integral is then over a surface at infinity and vanishes because g satisfies the radiation condition (see Section 3.3). It is appropriate to define an integral operator whose result $K[\psi]$ is a function of $\vec{x}$ and k,

$$K[\psi](\vec{x},k) = A(\vec{x},k)\int_\Upsilon [k^2\chi^2(\vec{x}_1,k) - \alpha(\vec{x}_1,k) - \beta(\vec{x}_1,\vec{x},k)]\psi(\vec{x}_1,k)\,d\Upsilon(\vec{x}_1) \quad (3.51)$$

so that (3.50) reduces to

$$\psi(\vec{x},k) = \psi_i(\vec{x},k) + K\psi(\vec{x},k). \quad (3.52)$$

which represents an integral formulation of the interaction between the field and the generalised constitutive parameter completely equivalent to the differential formulation (2.78). The integral formulation is a generalisation of the conventional volume-source [Jones 1964, Chapter 6] or polarisation-source principle [Bates and Ng 1972]. The scattered field is viewed as being comprised of radiations from sources, radiating according to some propagation principle, as determined by the particular choice of g. The conventional volume-source formulation adopts the free-space Green's function. The conventional equivalent sources, therefore, radiate isotropically. Refraction and diffraction are only accounted for implicitly by the need to treat (3.52) as an integral equation for the total field. The advantage of the generalised volume-source formulation over both the conventional volume-source formulation and the differential formulation is its capacity to instill desirable features into the expression for ψ by selection of the appropriate Green's function. The specific forms assumed by (3.50) for particular choices of Green's function, and hence K, are developed in Section 4.2.

Note that this argument is developed more fully in Appendix 3-A.

3.4.2 Surface-source formulation

Here, D is taken to be a region contained entirely within the free-space region. The free-space Green's function can be employed while the volume integral in (3.50) vanishes. Provided $\vec{x}$ is taken to lie within D,

$$\psi_s(\vec{x},k) = \int_S \left(\psi(\vec{x}_1,k)\frac{\partial g}{\partial n}(\vec{x},\vec{x}_1,k) - g(\vec{x},\vec{x}_1,k)\frac{\partial \psi}{\partial n}(\vec{x}_1,k) \right) dS(\vec{x}_1), \quad \vec{x} \in D. \quad (3.53)$$

The possibility exists, however, that x does not lie in D. The arguments leading to (3.50) remain the same except that the LHS is identically zero, since the δ-function has no effect :

$$-\psi_i(\vec{x},k) = \int_S \left(\psi(\vec{x}_1,k)\frac{\partial g}{\partial n}(\vec{x},\vec{x}_1,k) - g(\vec{x},\vec{x}_1,k)\frac{\partial \psi}{\partial n}(\vec{x}_1,k) \right) dS(\vec{x}_1), \quad \vec{x} \notin D. \quad (3.54)$$

The null field method of Section 4.3 is based on (3.53) and (3.54).

3.4.3 Volume-source formulation for discontinuous fields

The generalised volume-source formulation, of which most formulations of scattering employed in inverse scattering are special cases, is only valid provided the inhomogeneity is continuous everywhere. When χ is discontinuous across a number of interfaces, ψ and its first derivative exhibit jump discontinuities (see Section 2.5.4). In such cases, a more general form of Green's theorem [Shilov 1968, page 44] must be employed which incorporates terms to accommodate discontinuities of the field.

Discontinuities are often the dominant cause of scattering and refraction [Jones 1964, Chapter 6], [Beylkin 1985], [Robinson 1984]. It is consequently not always possible to disregard them or use smooth but rapidly varying approximations.

As mentioned in Section 2.5.4, the introduction of discontinuities poses a significant complication for inverse problems, where the location of the interfaces at which they occur may be unknown. Introduction of terms to account for discontinuities leads to nonlinear expressions for the field which cannot be inverted analytically. It is unlikely that a non-iterative exact inversion procedure can be found for such a relation. Efficient computational solution of direct problems incorporating discontinuities of the field is now becoming viable, so that promising prospects exist for solving the corresponding inverse problems by the procedures outlined in Section 6.7.

Computationally viable inverse diffraction algorithms such as diffraction tomography, discussed in Section 6.5, that are based on the generalised-volume source formulation do not explicitly account for the extra terms introduced by jump conditions. Such techniques are not completely inadequate, however, because the extra terms can be considered as distortions that are to be reduced by further processing of the image generated by relatively simple algorithms.

3-A GENERALISED VOLUME SOURCE FORMULATION OF DIFFRACTION AND A SEQUENCE OF APPROXIMATIONS

The following is a preprint of a paper (authors: D. G. H. Tan and R. H. T. Bates) to appear in *Wave Motion.*

Abstract

We introduce a new generalisation of the conventional volume source, or polarisation source, formulation of diffraction of scalar linear wave motion. The phase-integral, or Liouville-JWKB, principle is incorporated in such a way that a sequence of approximations is readily developed. The levels of the approximations, and the norms of the associated integral operators, are listed.

1 INTRODUCTION

Diffraction of linear scalar monochromatic wave motion by material bodies large compared to the wavelength is so computationally demanding that one is driven to invoke approximations. Provided all else is equal, the simpler the approximation that one can contrive the better, especially for inverse problems. The simplest approximation is undoubtedly that commonly named after Rayleigh and Gans [1, §6.13], sometimes called, particularly appropriately when the wave motion satisfies the Schrödinger equation [2] or the plasma wave equation [3], the Born approximation [4, §3.2].

The Rayleigh-Gans approximation can only be usefully invoked for a body whose largest linear dimension (expressed in wavelengths) is less than roughly $\pi/2$ divided by the average variation of the body's refractive index about the mean [5, §15]. Consequently the Rayleigh-Gans approximation is only effective in the real-world when either the refractive index is unity by definition (e.g. for non-relativistic quantum mechanics as characterised by the Schrödinger equation), or the body is extremely tenuous (as most crystals are to X-rays [6]), or one is merely interested in the gross statistics of diffracted fields (e.g. scattering of wave motion by randomly fluctuating media, such as sound waves in the ocean [7, §17.2], or radio waves in the troposphere [8]).

The conceptual and computational advantages of the Rayleigh-Gans approximation are so great, however, that one feels impelled to search for transformations of wave equations that will permit it to be usable in a wide range of practical acoustic and electromagnetic applications. The problem would be solved immediately if the macroscopic wave equation could be transformed into the form of the plasma wave equation (or Schrödinger equation). While this is easily done for wave motion depending only on a single spatial dimension [9],[10, §16], it appears to be impossible for wave motion exhibiting variations in two or more dimensions (despite recent claims to the contrary which are discussed in Section 3 below).

In this paper we derive an integral equation that reduces to a mere integral (in Rayleigh-Gans fashion) under conditions much wider than those for which the Rayleigh-Gans approximation is useful. The reason for this is that our formulation explicitly accounts for curvature of wavefronts and variable incremental phase delays.

The conventional volume source [1], or polarisation source [11] formulation is exact, but it only implicitly accounts for wavefront curvature, because its diffracted field comprises spherical waves (propagating as if in free-space) emanating from each point in space where the constitutive parameters (of the material body) differ from their free-space values. This tends to lead to numerical ill-conditioning and makes it difficult to introduce useful approximations.

Our new integral formulation constitutes a generalised volume source formulation, which is apparently superior to other extensions of the volume source principle [12] in that, besides accounting explicitly for refraction and variable phase delays, it also suggests several approximations, the error involved in each of which can be readily evaluated. Our treatment encompasses the phase-integral [13], or Liouville-JWKB [14], principle, together with the conventional volume source formulation, our new approximations of Rayleigh-Gans type, and some other approximations.

Section 2 states the diffraction problem. Ways of transforming the refractive index to unity are discussed in Section 3, in which we also discuss a generalised Green's function of Liouville-JWKB type. We introduce our generalised formulation and sequence of approximations in Sections 4 and 5 respectively. In Section 6 we assess the significance of our results and foreshadow their application to inverse problems.

2 THE DIFFRACTION PROBLEM

Three-dimensional space Υ, in which the position vector of an arbitrary point is denoted by $\vec{x}$, is partitioned according to

$$\Upsilon = \Upsilon_+ \cup \Upsilon_-, \tag{1}$$

where the volume of Υ_-, which is simply connected, is finite. Linear scalar wave motion (having the suppressed time-dependence $\exp(i\omega t)$ with ω being the angular frequency), which is characterised by the wave function $\psi = \psi(\vec{x}, k)$, where k is the free-space wavenumber, propagates throughout Υ according to

$$(\nabla^2 + k^2 n^2 + ika + b)\psi = -s, \tag{2}$$

where ∇^2 is the Laplacian operator, $n = n(\vec{x}, k)$ is the refractive index (varying in general with both $\vec{x}$ and k), $a = a(\vec{x})$ and $b = b(\vec{x})$ are the other (besides n) constitutive parameters of the medium supporting the wave motion, $i = \sqrt{-1}$, and $s = s(\vec{x}, k)$ is the source density, assumed confined to a region Υ_0 contained entirely within Υ_+, of the wave motion.

The wave function is expressed as

$$\psi = \theta + \phi, \tag{3}$$

where $\theta = \theta(\vec{x}, k)$ and $\phi = \phi(\vec{x}, k)$ are the incident and diffracted fields respectively. Space is free throughout Υ_+ (so that the source of the wave motion is outside the part of space, i.e. Υ_-, where the medium is inhomogeneous), i.e.

$$n = 1 \quad \text{and} \quad a = b = 0 \quad \text{for } \vec{x} \in \Upsilon_+. \tag{4}$$

The diffraction problem is: given n, a and b throughout Υ_-, and s throughout Υ_0, determine ϕ throughout Υ.

It is understood that the source density s gives rise to θ, which is to be thought of as impinging upon the medium. Diffraction by (or scattering from) the medium thereby generates ϕ.

3 SOME PRELIMINARY MANIPULATIONS OF THE WAVE EQUATION

A convenient "canonical" form for (2) is

$$(\nabla^2 + k^2 \nu^2 + \mu)\psi = -s, \tag{5}$$

where $\nu = \nu(\vec{x}, k)$ and $\mu = \mu(\vec{x})$ are defined by

$$\nu = n + ia/2kn \quad \text{and} \quad \mu = b + a^2/4n^2. \tag{6}$$

Restricting the spatial variation of ν (called the complex, or effective, refractive index), μ and ψ to a single Cartesian direction, described by the coordinate x, the wave equation (5) reduces to

$$d^2\psi/dx^2 + (k^2\nu^2 + \mu)\psi = -s. \tag{7}$$

On introducing the notation $q = d(\ln \nu)/d\tilde{x}$, $q' = dq/d\tilde{x}$, $\tilde{s}(\tilde{x}, k) = s(x, k)/\nu^{3/2}(x, k)$,

$$\tilde{x} = \int_0^x \nu(x, k)dx \quad \text{and} \quad \gamma = \gamma(\tilde{x}, k) = \nu^{1/2}(x, k)\psi(x, k), \tag{8}$$

we see that (5) transforms to [9],

$$d^2\gamma/d\tilde{x}^2 + k^2\gamma + (\mu/\nu^2 - q'/2 - q^2/4)\gamma = -\tilde{s}, \tag{9}$$

which is equivalent to the one-dimensional Schrödinger, or plasma wave, equation because ν has been removed from the term involving k^2 (i.e. the effective refractive index is now unity), provided the term in parentheses multiplying γ has no k-dependence. Strict equivalence with the Schrödinger equation requires that a be identically zero.

It has been suggested [15,16] that the above approach can be extended to two or more dimensions by introducing a curvilinear coordinate system $\vec{\tilde{x}}$ and defining its associated gradient operator by $\nabla = \nu\tilde{\nabla}$. We have since realised that this device is only approximate (and to a degree that it is difficult to assess) because such a curvilinear coordinate system attempts to coincide with the wavefronts and rays determined by the refractive index ν according to Fermat's principle. Each point in the medium, however, possess its own set of rays and hence requires its own coordinate system.

Since there is seemingly no general, exact way to transform the refractive index to unity, one has to seek alternative means of handling more-than-one-dimensional wave motion. For media characterised by certain pairs of constitutive parameters one can (as indicated in Section 4) construct Green's functions which exactly describe propagation from point sources of unit amplitude located arbitrarily in such media. On denoting such pairs of parameters by $\alpha = \alpha(\vec{x}, k)$ and $\beta = \beta(\vec{x}, \vec{\xi}, k)$, we write the equation satisfied by a generalised Green's function $g = g(\vec{x}, \vec{\xi}, k)$ as

$$(\nabla^2 + k^2\nu^2 + \alpha + \beta)g = -\left(1/B(\vec{\xi}, k)\right)\delta_3(\vec{x} - \vec{\xi}), \tag{10}$$

where the factor $B(\vec{\xi}, k)$ is included on RHS (10) to account for the variability of the medium's constitutive parameters, and where ∇^2 operates on the parts of g depending explicitly on $\vec{x}$. We evaluate $B(\vec{\xi}, k)$ by considering the strength of the Green's function singularity at $\vec{x} = \vec{\xi}$. It is worth noting that it is in general impossible to avoid including parameters (e.g. β and B) depending upon $\vec{\xi}$ as well as $\vec{x}$. Note also from (4), since space is free throughout Υ_+, that $g(\vec{x}, \vec{x}_0, k)$ is identical to the standard free space Green's function for all points for which the straight line from $\vec{\xi}$ to $\vec{x}$ does not intersect Υ_-, implying that $B(\vec{\xi}, k)$ can be normalised to unity whenever $\vec{\xi} \in \Upsilon_+$. In particular, $B(\vec{\xi}, k) = 1$ whenever $\vec{\xi} \in \Upsilon_0$. As specified in Section 2, we define θ by

$$\theta(\vec{x}, k) = \int_{\Upsilon_0} s(\vec{\xi}, k)g(\vec{x}, \vec{\xi}, k)\, d\Upsilon(\vec{\xi}), \tag{11}$$

where $d\Upsilon(\vec{\xi})$ is the element of volume. On multiplying (10) by $s(\vec{\xi}, k)$ and integrating over Υ_0, we see from (11) that

$$(\nabla^2 + k^2\nu^2 + \alpha)\theta(\vec{x}, k) = -s_0(\vec{x}, k) - s(\vec{x}, k), \tag{12}$$

$$s_0 = s_0(\vec{x}, k) = \int_{\Upsilon_0} s(\vec{x}_0, k)\beta(\vec{x}, \vec{x}_0, k)g(\vec{x}, \vec{x}_0, k)\, d\Upsilon(\vec{x}_0). \tag{13}$$

The most obvious way to invoke g to define ϕ, but consistently with (2) and (3), is to introduce an equivalent source density, say $\tau = \tau(\vec{x}, k)$, and to write

$$\phi(\vec{x}, k) = \int_{\Upsilon_-} \tau(\vec{\xi}, k)g(\vec{x}, \vec{\xi}, k)\, d\Upsilon(\vec{\xi}). \tag{14}$$

Combining (10) and (14) gives

$$(\nabla^2 + k^2\nu^2 + \alpha)\phi + \tau_- + \tau/B = 0, \tag{15}$$

$$\tau_- = \tau_-(\vec{x}, k) = \int_{\Upsilon_-} \tau(\vec{\xi}, k)\beta(\vec{x}, \vec{\xi}, k)g(\vec{x}, \vec{\xi}, k)\, d\Upsilon(\vec{\xi}). \tag{16}$$

It follows from (3),(5), and (11) through (16) that

$$\tau(\vec{x},k)/B(\vec{x},k) = [\mu(\vec{x})-\alpha(\vec{x},k)]\,\theta(\vec{x},k) - s_0(\vec{x},k) + \int_{\Upsilon_-}\left[\mu(\vec{x})-\alpha(\vec{x},k)-\beta(\vec{x},\vec{\xi},k)\right]\tau(\vec{\xi},k)g(\vec{x},\vec{\xi},k)\,d\Upsilon(\vec{\xi}), \quad (17)$$

which, for $\vec{x}\in\Upsilon_-$, is to be thought of as a Fredholm integral equation of the second kind for $\tau(\vec{x},k)$. This is the essence of the approach introduced by Hudimac [12].

The approximation of Rayleigh-Gans type to (17) is obtained by neglecting the integral over Υ_-. An iterative solution for τ can be constructed be defining $\tau = \Sigma_{m=0}^{\infty}\tau_m$ where

$$\tau_{m+1}(\vec{x},k) = \int_{\Upsilon_-} h(\vec{x},\vec{\xi},k)\tau_m(\vec{\xi},k)\,d\Upsilon(\vec{\xi}), \quad (18)$$

with $h(\vec{x},\vec{\xi},k) = \left(\mu(\vec{x})-\alpha(\vec{x},k)-\beta(\vec{x},\vec{\xi},k)\right)g(\vec{x},\vec{\xi},k)$ and $\tau_0(\vec{x},k) = h(\vec{x},\vec{x}_0,k)$. It is not easy to make *a priori* estimates for the convergence of the series for τ. Neither can one readily assess the errors involved in neglecting τ_m for m greater than any particular integer.

4 THE GENERALISED VOLUME SOURCE FORMULATION

We adopt the definitions (10) and (11) for the Green's function and incident field, so that we hold to (12), but we define ϕ differently. We note from (3), (5), and (12) that

$$(\nabla^2 + k^2\nu^2 + \alpha)\phi = -\sigma\psi + s_0, \quad (19)$$

where the negative of RHS (19) is appropriately called the generalised volume source density, with

$$\sigma = \sigma(\vec{x},k) = \mu(\vec{x}) - \alpha(\vec{x},k). \quad (20)$$

Taking both g and ϕ to satisfy the Sommerfeld radiation condition at infinity [17], which is compatible with (10) and (19), and assuming the field and its normal derivative are continuous on the boundary of Υ_- and Υ_+, it follows from the divergence theorem [18] that

$$0 = \int_{\Upsilon_-}\left[g(\vec{x},\vec{\xi},k)\nabla^2\phi(\vec{x},k) - \phi(\vec{x},k)\nabla^2 g(\vec{x},\vec{\xi},k)\right]d\Upsilon(\vec{x}), \quad (21)$$

which leads, when (10) and (19) are invoked, with $\vec{x}$ and $\vec{\xi}$ being interchanged, to

$$\phi(\vec{x},k) = \varpi(\vec{x},k) + B(\vec{x},k)\int_{\Upsilon_-}\rho(\vec{x},\vec{\xi},k)\phi(\vec{\xi},k)\,d\Upsilon(\vec{\xi}), \quad (22)$$

$$\varpi = \varpi(\vec{x},k) = B(\vec{x},k)\int_{\Upsilon_-}[\sigma(\vec{\xi},k)\theta(\vec{\xi},k) - s_0(\vec{\xi},k)]g(\vec{\xi},\vec{x},k)\,d\Upsilon(\vec{\xi}), \quad (23)$$

$$\rho = \rho(\vec{x},\vec{\xi},k) = [\sigma(\vec{\xi},k) - \beta(\vec{\xi},\vec{x},k)]g(\vec{\xi},\vec{x},k). \quad (24)$$

Eqn. (22) is our generalised volume source formulation for the diffracted field. It is instructive to rewrite it as a Fredholm integral equation of the second kind [19] for the unknown ψ, in terms of the known θ and the known constitutive parameters, e.g. for $\vec{x}\in\Upsilon_-$,

$$\psi(\vec{x},k) = \theta(\vec{x},k) + \varpi_0(\vec{x},k) + B(\vec{x},k)\int_{\Upsilon_-}\rho(\vec{x},\vec{\xi},k)\psi(\vec{\xi},k)\,d\Upsilon(\vec{\xi}), \quad (25)$$

$$\varpi_0(\vec{x},k) = B(\vec{x},k)\int_{\Upsilon_-}[\beta(\vec{\xi},\vec{x},k)\theta(\vec{\xi},k) - s_0(\vec{\xi},k)]g(\vec{\xi},\vec{x},k)\,d\Upsilon(\vec{\xi}). \quad (26)$$

Once ψ is evaluated, ϕ is found from (3). Since the condition number [20] of the matrix inverse which must be computed in order to evaluate ψ tends to worsen appreciably as k increases, and the number of terms required to represent ψ is linearly related to frequency for each of the three spatial dimensions, so that the computational cost is proportional to $k^6 \log k^3$, it is generally necessary to resort to iteration :

$$\psi(\vec{x},k) = \sum_{n=0}^{\infty} \phi_n(\vec{x},k) \quad \text{with} \quad \phi_0(\vec{x},k) = \theta(\vec{x},k) + \varpi_0(\vec{x},k), \tag{27}$$

so that $\phi(\vec{x},k) = \varpi_0(\vec{x},k) + \sum_{n=0}^{\infty} \phi_n(\vec{x},k)$. Note, for $n > 0$, that

$$\phi_n(\vec{x},k) = B(\vec{x},k) \int_{\Upsilon_-} \rho(\vec{x},\vec{\xi},k)\phi_{n-1}(\vec{\xi},k)\, d\Upsilon(\vec{\xi}), \tag{28}$$

represents n-th order multiple scattering within the inhomogeneous medium.

The Neumann series (27) is only useful if it converges, which requires $\mathcal{N} < 1$ where the norm $\mathcal{N}$ is given by

$$\mathcal{N} = \sup_{\|\phi\|=1} \left\| B(\vec{x},k) \int_{\Upsilon_-} \rho(\vec{x},\vec{\xi},k)\phi(\vec{\xi},k)\, d\Upsilon(\vec{\xi}) \right\|, \tag{29}$$

computed in any convenient norm. We adopt the "uniform norm", $\|\cdot\|_\infty$, given by

$$\|\phi\|_\infty = \sup_{\vec{x}\in\Upsilon_-} |\phi(\vec{x},k)|. \tag{30}$$

A useful feature of our generalised volume source formulation is that it permits some control over $\mathcal{N}$ by appropriate choice of Green's function. The form of the latter governs the magnitude of $\mathcal{N}$ and determines its frequency dependence. Another advantage is that approximate formulations are readily devised, and their accuracy can be estimated straightforwardly, as we demonstrate in Section 5.

4.1 Conventional Volume Source Formulation

The standard free-space Green's function is adopted [1], [21, §7.2] :

$$g(\vec{x},\vec{\xi},k) = (1/4\pi|\vec{x}-\vec{\xi}|)\exp(-ik|\vec{x}-\vec{\xi}|), \tag{31}$$

so that

$$\alpha = k^2(1-\nu^2), \quad \beta = \varpi_0 = 0, \quad B = 1, \quad \text{and} \quad \sigma = k^2(\nu^2-1) + \mu. \tag{32}$$

The integral equation (22) becomes formally simple but tends to be computationally inefficient because σ is usually appreciable throughout Υ_-. Furthermore, $\mathcal{N} = O(k^2)$ implying that this conventional formulation is only effective when the wavelength is below a limit set by the physical dimensions of, and the magnitude of fluctuations of $(\nu^2 - 1)$ within, Υ_-. This is intuitively satisfying because the formulation does not accommodate incremental delays and curvatures of wavefronts, which are always manifested by wave motion in inhomogeneous media.

4.2 Liouville-JWKB Volume Source Formulation

The Green's function is chosen to be

$$g(\vec{x},\vec{\xi},k) = (\nu^{-1/2}/4\pi S)\exp(-ikS), \tag{33}$$

with $S = S(\vec{x},\vec{\xi},k)$ satisfying the eikonal [22,23]

$$\nabla S \cdot \nabla S = \nu^2, \tag{34}$$

implying that

$$S = \int_{l(\vec{\xi})}^{l(\vec{x})} \nu \, dl, \tag{35}$$

where $l = l(\vec{x})$ describes path length along the concourse of rays from $\vec{\xi}$ to all points $\vec{x}$ through which the rays pass. On defining

$$2p = \ln(\nu), \quad q = \ln(S) \quad \text{and} \quad z = \nabla^2 q - \nabla q \cdot \nabla q - 2\nabla p \cdot \nabla q, \tag{36}$$

we see that

$$\alpha = \nabla^2 p - \nabla p \cdot \nabla p, \quad \beta = (1 + ikS)z, \quad \text{and} \quad B = \nu^{-3/2}, \tag{37}$$

indicating that α is only appreciable where the constitutive parameters of the medium change significantly. This, together with the $O(k)$ dependence of β at high frequencies, tends to reduce $\mathcal{N}$ markedly by comparison with the conventional formulation. This stems from g satisfying the mathematical physics of the problem (i.e. incremental delays and curvatures of wavefronts are explicitly accommodated).

4.3 Forward Scattering Liouville-JWKB Volume Source Formulation

When the sources of the incident field are sufficiently far away from the inhomogeneity, and the medium is tenuous enough, that scattering is predominantly in the forward direction (as is conventionally assumed when the standard parabolic approximation [24] is invoked), it can be appropriate to choose the Green's function

$$g(\vec{x},\vec{\xi},k) = \nu^{-1/2}\exp(-ikS), \tag{38}$$

with (34) through (37) still applying, except that

$$\beta = ikSz, \tag{39}$$

where z is defined in (36). The usefulness of this Green's function relies on the ability to describe the field in terms of plane waves. The norm $\mathcal{N}$ exhibits the same general behaviour as for Section 4.2, but its magnitude is significantly reduced when certain physically interesting conditions apply, as is demonstrated in Section 5.4.

5 VARIOUS LIOUVILLE-JWKB APPROXIMATIONS

Before the formulations introduced in Sections 4.2 and 4.3 can be implemented, the phase function $S(\vec{x},\vec{\xi},k)$ (it might be better called the wavefront function) must be evaluated throughout Υ_- (i.e. for all $\vec{x}$) for each $\vec{\xi} \in \Upsilon_-$. Only a finite number of points, say M, can of course be considered in practice. It follows that whatever integral

formulas — e.g. (14) or (22) — are invoked for ϕ, the latter can only be evaluated in practice by approximating an integral by a summation. In order to faithfully reproduce the interference of the fields scattered from the interior of Υ_-, it must be partitioned into sub-regions having linear dimensions $\lambda/\ell\bar{\nu}$ rather less than the average wavelength $\lambda/\bar{\nu}$ within Υ_-, where $\bar{\nu}$ is the average value of ν and ℓ is not less than 2. If $N\lambda$ is the mean linear dimension of Υ_-, we see that a useful estimate of M is $(\ell\bar{\nu}N)^3$. This estimate can be reduced when (as noted in Section 4.2) variations of the generalised volume source density is small in parts of Υ_-.

Except when the simplification noted in Section 5.2 can be justified, ray-tracing algorithms [25,26] must be invoked to compute $S(\vec{x},\vec{\xi},k)$ for the point $\vec{\xi}$ at the centre of each of the abovementioned sub-regions.

We now introduce a number of approximations to the formulations introduced in Sections 4.2 and 4.3.

5.1 Far-field approximation to θ

This applies when the separation of all $\vec{x}_0 \in \Upsilon_0$ from all $\vec{x} \in \Upsilon_-$ is large enough that $\beta(\vec{x},\vec{x}_0,k)$ is everywhere negligible. It then transpires that

$$s_0 = \varpi_0 = 0 \quad \text{and} \quad \phi_0(\vec{x},k) = \theta(\vec{x},k). \tag{40}$$

The values of $|\vec{x}-\vec{x}_0|$ for which this approximation is valid, to any required accuracy, depends upon the behaviour within Υ_- of p and q (refer to Sections 5.3 and 5.4 below).

5.2 Approximate evaluation of $S(\vec{x},\vec{\xi},k)$

On introducing the notation

$$r = |\vec{x}-\vec{\xi}|, \quad S = \bar{\nu}(r+\zeta) \quad \text{and} \quad \nu = \bar{\nu}(1+\epsilon), \tag{41}$$

where $\zeta = \zeta(\vec{x},\vec{\xi},k)$ and $\epsilon = \epsilon(\vec{x},k)$, we see that

$$\partial\zeta/\partial r = \epsilon, \tag{42}$$

satisfies (34) to first order, assuming that $|\epsilon|$ is appreciably less than unity throughout most of Υ_-, which is to be expected in many important practical applications (e.g. for light and radio waves propagating though planetary and stellar atmospheres, sonar in the ocean, medical-diagnostic ultrasound). If χ is the angle made by ∇S with the $(\vec{x}-\vec{\xi})$ direction at $\vec{x}$, then

$$\left|\frac{\partial\zeta/\partial t}{\partial\zeta/\partial r}\right| = \tan\chi, \tag{43}$$

where the t-axis intersects the point $\vec{x}$ perpendicularly to $(\vec{x}-\vec{\xi})$, such that there is no component of ∇S perpendicular to both $(\vec{x}-\vec{\xi})$ and the t-axis. Now, substituting (41) and (43) into (34), and solving for the positive root of the resulting quadratic equation for $\partial\zeta/\partial r$, gives

$$\partial\zeta/\partial r = \cos^2\chi\left[\left(1+\frac{2\epsilon+\epsilon^2}{\cos^2\chi}\right)^{1/2}-1\right]. \tag{44}$$

Provided ϵ is moderately small in the abovementioned sense, implying that χ is unlikely to exceed, say, 0.5, we see that $\partial\zeta/\partial r$ is approximately $(\epsilon-(\epsilon^2/2)\tan^2\chi)$. It follows that the relative error in the estimate (42) for $\partial\zeta/\partial r$ is roughly $\epsilon\chi^2/2$, which is unlikely

to exceed 0.1 in any of the applications mentioned above, and can be expected to be much less. The formula (42) is thus likely to be acceptably accurate under wide sets of circumstances.

When the approximation (42) is valid, S can be evaluated much more conveniently than by the computationally expensive procedure (which must incorporate comprehensive error-checking) of ray-tracing. We evaluate ζ simply by integration of ϵ along radial paths between all points $\vec{\xi}$ and $\vec{x}$ in Υ_-.

5.3 Approximate evaluation of α and β

Substituting (36) and (41) into (37) gives

$$\alpha = \frac{\nabla^2\epsilon}{2(1+\epsilon)} - \frac{3\nabla\epsilon\cdot\nabla\epsilon}{4(1+\epsilon)^2}, \tag{45}$$

$$z = \frac{2/r+\nabla^2\zeta}{r+\zeta} - \frac{2+4\partial\zeta/\partial r+2\nabla\zeta\cdot\nabla\zeta}{(r+\zeta)^2} - \frac{\partial\epsilon/\partial r+\nabla\epsilon\cdot\nabla\zeta}{(1+\epsilon)(r+\zeta)}. \tag{46}$$

Provided $|\epsilon|$ and $|\nabla\epsilon\cdot\nabla\epsilon/(1+\epsilon)\nabla^2\epsilon|$ are both somewhat less than unity throughout most of Υ_-, which one expects to be true in many (perhaps the great majority of) applications of physical interest, it is justifiable to approximate (45) by

$$\alpha = \nabla^2\epsilon/2 = (1/2\bar{\nu})\nabla^2\nu. \tag{47}$$

The behaviour of z needs to be examined both near $r = 0$ and for r large. Since, as noted in the introductory paragraph to Section 5, the spatial variation of ν is negligible on the average for r less than some r_0, we see from (41), (42), (44), and (46) that

$$z = 0 \quad \text{for} \quad r < r_0. \tag{48}$$

Consequently integrals, such as those on RHS (25), are finite. The lower limit of the integration with respect to r is r_0, so that the integrand remains bounded.

Note from (37) that β behaves the same as z for small r.

Reverting to the approximation (42) for $\partial\zeta/\partial r$, we see that (46) reduces for large r to

$$z \approx -\frac{2\epsilon}{r^2} + \frac{\partial^2\zeta/\partial t^2 - (\partial\epsilon/\partial r)(\partial\zeta/\partial r)}{r} + \frac{2\zeta}{r^3} + \cdots, \tag{49}$$

on the assumption that the angle χ introduced in (43) is appreciably less than unity (i.e. is as small as one expects in the applications mentioned in Section 5.2). Whether, in any particular instance, the greater contribution comes from the $1/r$ or the $1/r^2$ term in (49) depends upon the magnitudes of the spatial derivatives of ϵ compared with $|\epsilon|$ throughout Υ_-. The maximum absolute value of z depends, therefore, on (48) and on the variations mentioned in the previous sentence, and so is appropriately expressed as

$$|z| < Z/r_0, \tag{50}$$

where Z is the maximum value of rz.

Note from (37) that β varies as $O(zr)$ for large r.

5.4 Approximate evaluation of β for the forward scattering volume source formulation

Reference to Section 4.3 indicates that the radial coordinate r introduced in (41) should be replaced by a Cartesian coordinate, x say, parallel to the forward scattering

direction. The approximate formula (47) for α still applies, but the factor $(r + \zeta)$ appearing in three denominators in (46) should be replaced by x, because Section 4.3 is (in effect) predicated upon the sources of the wave motion being far removed from Υ_- (refer to Section 5.1). It follows from (37), (39) and (49) that

$$\beta \approx ik\bar{\nu}[\partial^2\zeta/\partial y^2 - (\partial\epsilon/\partial x)(\partial\zeta/\partial x)], \tag{51}$$

where it is appropriate to replace t by the explicitly Cartesian coordinate y.

5.5 Liouville-JWKB Rayleigh-Gans/Born approximation

By analogy with the conventional Rayleigh-Gans approximation [1, §6.13] we approximate (27) by

$$\psi(\vec{x}, k) = \phi_0(\vec{x}, k) + \phi_1(\vec{x}, k), \tag{52}$$

which of course assumes that $\mathcal{N} \ll 1$. The considerable attraction of this approximation, which has in a limited sense been employed by Clayton and Stolt [27], is that the integral equation (25) for $\psi(\vec{x}, k)$ is reduced to a mere integral.

5.6 Norms of integral operators

When Rayleigh-Gans approximations are inadequately accurate, one must resort either to some kind of matrix inversion of (22) or to iteration. As noted in Section 4, the former is often impracticable. The magnitude of $\mathcal{N}$ as defined by (29) and (24), governs first of all whether the Neumann series in (27) converges, and secondly how many iterations are required to solve the diffraction problem to any desired accuracy. If $\mathcal{N}$ is greater than unity, (27) diverges. At the other extreme, if $\mathcal{N}$ is appreciably less than unity, then it may be sufficient to sum just the first two terms of the Neumann series, resulting in the approximation (52). It is instructive to examine the values at high frequencies (i.e. k is large) of the norms arising from the Green's functions of Section 4 in terms of $\hat{N}\lambda$, the largest linear dimension of Υ_- expressed in wavelengths.

In view of (32), the norm of the operator arising from the conventional Green's function is

$$\mathcal{N} = \|\mu + k^2(\nu^2 - 1)\|_\infty(\hat{N}\lambda)^2/2 \tag{53}$$

$$\approx 20\|(\nu^2 - 1)\|_\infty\hat{N}^2 \quad \text{as} \quad k \to \infty. \tag{54}$$

In contrast, the norm for the Liouville-JWKB Green's function is

$$\mathcal{N} = Z\hat{N}, \tag{55}$$

where Z is defined by (50). These expressions have an interesting interpretation in the context of inverse diffraction where, in order to obtain high resolution in reconstructions, it is desirable to employ high frequency radiation. The norms stated above both increase with increasing $\hat{N}$, which is itself proportional to k. For the Liouville-JWKB volume source formulation, $\mathcal{N}$ is, however, only linearly proportional to $\hat{N}$, so the Liouville-JWKB formulation can be expected to be valid for a larger range of frequencies than the conventional formulation.

6 CONCLUSIONS

We have placed the Rayleigh-Gans/Born approximation into a physically acceptable context (i.e. one explicitly accommodating wavefront curvature and differential phase

delays) by expressing it through the phase-integral or Liouville-JWKB principle. While in a sense merely following a long line of predecessors, we claim that the new slant introduced in Section 4 gives better insight into the diffraction problem and makes it easier to devise and assess the level of various approximations (as done in Section 5). In fact, when it is justifiable to approximate S as in Section 5.2, the Liouville-JWKB version (Section 5.5) of the Rayleigh-Gans approximation is only somewhat less convenient than the conventional version.

The approximations discussed in Section 5.2 are reasonable from the points of view of perhaps the great majority of applications of linear wave motion that it may be worth attempting to evaluate them comprehensively in practical contexts. One might then discover why, for instance, simple imaging techniques function surprisingly well under unlikely conditions, e.g. ultrasonic imaging through severely distorting media [28].

Our chief motivation for devising the generalised volume source formulation has been to find a formalism for the direct diffraction problem which can be conveniently incorporated into the Newton-Kantorovich formulation of the inverse problem. The conventional Rayleigh-Gans formulation is highly convenient in this regard, as is well known [29,30]. We find that the approximations introduced in Section 5.2 and 5.3, when combined with the Liouville-JWKB Rayleigh-Gans approximation of Section 5.5, lead to what seems to be a widely applicable inverse formalism [31].

7 ACKNOWLEDGEMENTS

One of us (DGHT) acknowledges the award of a Postgraduate Research Scholarship from the New Zealand University Grants Committee and an Australian Guaranteed Corporation (NZ) Ltd. Young Achievers Award 1987.

References

[1] D. S. Jones, *The Theory of Electromagnetism*, Pergamon Press, New York (1964).

[2] L. I. Schiff, *Quantum Mechanics*, McGraw-Hill, New York, second edition (1955). International series in Pure and Applied Physics.

[3] B. DeFacio and J. H. Rose, "Inverse-scattering theory for the non-spherically-symmetric three-dimensional plasma wave equation", *Phys. Rev. A*, 31, 897–902 (1985).

[4] R. G. Newton, *Scattering Theory of Waves and Particles*, McGraw-Hill, New York (1966).

[5] L. A. Chernov, *Wave Propagation in a Random Medium*, McGraw-Hill, New York (1960). Translated from Russian by R. A. Silverman.

[6] G. N. Ramachandran and R. Srinivasan, *Fourier Methods in Crystallography*, Wiley-Interscience, New York (1970).

[7] A. Ishimaru, *Wave Propagation and Scattering in Random Media*, Volume 2, Academic Press, New York (1978).

[8] A. D. Wheelon, "Radio-wave scattering by tropospheric irregularities", *J. Res. (U.S.) Natn Bur. Standards*, 63D, 205–232 (1959).

[9] R. H. T. Bates and D. J. N. Wall, "Chandrasekhar transformations improve convergence of computations of scattering from linearly stratified media", *IEEE Trans. Antennas Propagat.*, AP-24, 251–253 (1976).

[10] L. M. Brekhovskikh, *Waves in Layered Media*, Academic Press, New York (1960).

[11] R. H. T. Bates and F. L. Ng, "Polarisation-source formulation of electromagnetism and dielectric-loaded waveguides", *Proc. Instn. elect. Engrs*, 119, 1568–1574 (1972).

[12] A. A. Hudimac, "Diffraction by iterative ray-theory techniques", *J. acoust. Soc. Am.*, 56, 1342–1346 (1974).

[13] J. Heading, *An Introduction to Phase-integral Methods*, John Wiley and sons, New York (1962).

[14] B. Ursin, "Seismic migration using the WKB approximation", *Geophys. J. R. astr. Soc.*, 79, 339–352 (1984).

[15] R. H. T. Bates and R. P. Millane, "Time domain approach to inverse scattering", *IEEE Trans. Antennas Propagat.*, AP-29, 359–363 (1981). Special Issue on Inverse Methods on Electromagnetics.

[16] R. H. T. Bates, "Full-wave computed tomography. Part 1: Fundamental theory", *IEE Proceedings Part A*, A131, 610–615 (1984).

[17] C. Müller, *Foundations of the Mathematical Theory of Electromagnetic Waves*, Springer-Verlag, New York (1969).

[18] F. John, *Partial Differential Equations*, Springer-Verlag, New York, third edition (1978).

[19] P. P. Zabreyko et al., *Integral equations — a reference text*, Noordhoff, Leyden (1975). Translated and edited by T. O. Shaposhnikova and R. S. Anderssen and S. G. Mikhlin.

[20] R. F. Harrington, *Field Computation by Moment Methods*, MacMillan, New York (1968).

[21] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics*, McGraw-Hill, New York (1953). Volume 1 : Chapters 1 to 8; volume 2 : Chapters 9 to 13.

[22] M. Born and E. Wolf, *Principles of Optics*, Pergamon Press, fourth edition (1970).

[23] G. Ghione, I. Montrosset, and L. B. Felsen, "Complex ray analysis of radiation from large apertures with tapered illumination", *IEEE Trans. Antennas Propagat.*, AP-29, 684–693 (1984).

[24] J. A. DeSanto, "Relation between the solutions of the Helmholtz and parabolic equations for sound propagation", *J. acoust. Soc. Am.*, 62, 295–297 (1977).

[25] J. Fawcett and H. B. Keller, "Three-dimensional ray tracing and geophysical inversion in layered media", *SIAM J. Appl. Math.*, 45, 492–501 (1985).

[26] M. H. Lawrence, "Ray theory modeling applied to low-frequency acoustic interaction with horizontally stratified ocean bottoms", *J. acoust. Soc. Am.*, 78, 649–658 (1985).

[27] R. W. Clayton and R. H. Stolt, "A Born-WKBJ inversion method for acoustic reflection data", *Geophysics*, 46, 1559–1567 (1981).

[28] F. S. Foster, M. S. Patterson, M. Arditi, and J. W. Hunt, "The conical scanner: a two transducer ultrasound scatter imaging technique", *Ultrason. Imaging*, 3, 62–82 (1981).

[29] E. Wolf, "Three-dimensional structure determination of semi-transparent objects from holographic data", *Opt. Commun.*, 1, 153–156 (1969).

[30] W. H. Carter, "Inverse scattering in the first Born approximation", *Opt. Eng.*, 23, 204–209 (1984).

[31] D. G. H. Tan, *Implementable Multi-dimensional Inverse Scattering Theory*, PhD thesis, Department of Electrical and Electronic Engineering, University of Canterbury, Christchurch, New Zealand (1988).

Chapter 4

APPROACHES TO THE DIRECT PROBLEM

4.1 INTRODUCTION

This chapter is concerned with the second task involved in examining an inverse problem, namely, analysing the corresponding direct problem. The starting point for the direct problem is the canonical form (2.78) of the underlying frequency-domain equation (2.71), or its equivalent integral formulations derived in Section 3.4. In the direct problem, the generalised constitutive parameter χ is known throughout the region Υ, as is the incident field ψ_i. The canonical frequency-domain equation then becomes a conventional linear partial differential equation to be solved for the unknown function ψ, and the integral formulations become linear Fredholm integral equations of the second kind.

Solution of the direct problem defines a mapping (see Section 1.4)

$$\Omega : C \rightarrow E, \tag{4.1}$$

from the space C of generalised constitutive parameters and incident fields, to the space E characterising the scattered fields. When the direct problem is considered in isolation, accurate numerical determination of this mapping for a particular element of C is performed by resorting to computational solution of either a differential or integral equation (see, for example, Harrington [1968] or Smith [1978]). A thorough discussion of the direct problem requires several theses in its own right, as is attested to by the amount of active research interest it continues to generate. It must be remembered that, for this thesis, examination of a direct problem is merely a prerequisite to solving an inverse problem. It is preferable (although not always possible) not to resort to numerical solution of the differential or integral equations defining the direct problem, but rather to merely analytically establish integral expressions for the scattered field ψ_s (or, equivalently, the total field ψ) in terms of the generalised constitutive parameter and incident field. Of particular interest are relations that can be inverted to give the generalised constitutive parameter in terms of the incident and scattered fields. This chapter, therefore, contains only those aspects of the direct problem that bear usefully upon the inverse scattering techniques discussed in the next chapter.

Although the interaction between the emanations and the generalised constitutive parameter is stated concisely in both the differential and integral formulations

of scattering, a general explicit expression for ψ in terms of χ or, of more relevance to inverse problems, an expression for χ in terms of ψ, cannot be stated. This is because the direct problem for a given generalised constitutive parameter χ is linear with respect to different incident fields, but depends in a nonlinear fashion on the generalised constitutive parameter. Solutions to inverse problems are considerably simpler (both conceptually and numerically) if the (unique) solution of the direct problem can be usefully approximated by an expression for (some simple functional of) the scattered field that is linear with respect to χ. Consequently, approximate solutions of the direct problem are often employed to facilitate tractable inversion algorithms. Section 4.2 employs the generalised volume-source formulation (which is introduced in Section 3.4.1) with the view to deriving approximate expressions for the scattered field whose validity can be readily assessed. The most convenient analytical simplification is the neglect of multiple scattering.

Subsections of Section 4.2 contain the two major approximate solutions of the direct problem that are employed in inverse scattering problems. The first is the Born approximation [Newton 1966, Chapter 3] (also known as the Rayleigh-Gans approximation [Jones 1964, Chapter 6]). The second is the Rytov approximation [Mueller 1980]. A key feature in the derivation of the Rytov approximation is the transformation of (2.78), which is a linear partial differential equation, to a nonlinear partial differential equation. The far-field forms of the Born and Rytov approximations are shown to be particularly convenient for numerical solution of direct problems, the corresponding inverse problem being considered in Section 6.5. Appendix 3-A contains a sequence of approximate solutions to the direct problem derived from the generalised volume-source formulation of diffraction.

There are many occasions, if they do not constitute the majority, where simple analytical expressions for the scattered field are elusive, and numerical methods, evaluable to arbitrary accuracy (in principle), for solving the direct problem must be invoked. One such method is the null-field method presented in Section 4.3, and is based on the surface formulation of Section 3.4. The computational procedure is invoked in a particular method of solution of an associated inverse problem in Appendix 6-B.

The major computational effort involved in the null-field method is the determination of an equivalent source distribution that produces the same scattered field as the true obstacle. The physical optics approximation outlined in Section 4.4 removes this difficulty while retaining sufficient accuracy to be usefully applied in many applications.

4.2 GENERAL VOLUME-SOURCE FORMULATIONS

When invoking the generalised volume-source formulation of scattering to describe the direct problem, the incident field ψ_i and the integral operator K, as defined by (3.48) and (3.51) respectively, are both known, so that (3.52) is a linear Fredholm integral equation of the second kind [Zabreyko *et al.* 1975, Chapter 1]. It is worth noting that the values of the field required to compute $K\psi$ are unavailable in the inverse problem, because the field is not measured within the region containing the inhomogeneity, over which the integral is performed. The Neumann series [Zabreyko *et al.* 1975, Chapter 2]

is formed by defining

$$\begin{aligned} \psi^{(0)}(\vec{x},k) &= \psi_i(\vec{x},k), && (4.2)\\ \psi^{(n)}(\vec{x},k) &= K[\psi^{(n-1)}](\vec{x},k) \\ &= K^n\psi^{(0)}, \quad n \geq 1, && (4.3) \end{aligned}$$

so that

$$\psi = \lim_{n\to\infty} \sum_{i=0}^{n} \psi^{(i)} = \psi_i + \lim_{n\to\infty} \sum_{i=0}^{n-1} K\psi^{(i)}, \tag{4.4}$$

is the (unique) solution to the direct problem, provided the series converges. Convergence of the Neumann series is ensured by the (sufficient, but not necessary) condition $\|K\| < 1$, measured in some appropriate norm, which depends on the Green's function that determines K. The operator K depends on the generalised constitutive parameter χ, so that terms defined by (4.3) for $n \geq 2$ are nonlinear with respect to χ. Such nonlinear terms may be neglected if the convergence criterion is replaced by the stronger restriction $\|K\| \ll 1$. Specific instances when the resulting linear approximation is valid are now discussed. The maximum linear dimension of the region Υ_s is denoted by L.

4.2.1 Born approximations

When the conventional Green's function (see Section 3.3) is employed, the integral operator K becomes

$$\left[K^{(B)}\psi\right](\vec{x},k) = \int_{\Upsilon_s} k^2 \left[\chi^2(\vec{x}_1,k) - 1\right] \psi(\vec{x}_1,k) g(\vec{x},\vec{x}_1,k)\, d\Upsilon_s(\vec{x}_1). \tag{4.5}$$

The restriction on the norm of this operator to ensure convergence of the series (4.4) can be written [Leeman *et al.* 1985]

$$k^2 \max_{\vec{x}\in\Upsilon_s} |\chi^2(\vec{x},k) - 1| L^3 < 1. \tag{4.6}$$

The terms defined by (4.3), which represent multiple scattering, assume a significantly simpler form when $\vec{x} = R\hat{x}$ (a caret adorns unit vectors) is in the far-field of Υ_s, so that the free-space Green's function is adequately approximated by (3.27) :

$$\psi^{(B,N)}(\vec{x},k) = \int_{\Upsilon_s} \left[\chi^2(\vec{x}_1,k) - 1\right] \psi^{(B,N-1)}(\vec{x}_1,k) e^{ik\hat{x}\cdot\vec{x}_1} d\Upsilon_s(\vec{x}_1), \tag{4.7}$$

where a factor depending only on R and k has been neglected. Since χ assumes the value unity outside Υ_s, the integral can be extended to the whole of Υ. Invocation of (2.109) simplifies (4.7) to

$$\psi^{(B,N)}(\vec{x},k) = \mathcal{F}[(\chi^2 - 1)\psi^{(B,N-1)}](\lambda^{-1}\hat{x},k), \tag{4.8}$$

where $\lambda = 2\pi/k$ is the wavelength of the radiation. That is, a Fourier transform relationship exists between $\psi^{(B,N)}$ and $(\chi^2 - 1)\psi^{(B,N-1)}$. The Fourier transform relation still holds when χ is frequency dependent, provided k is kept constant.

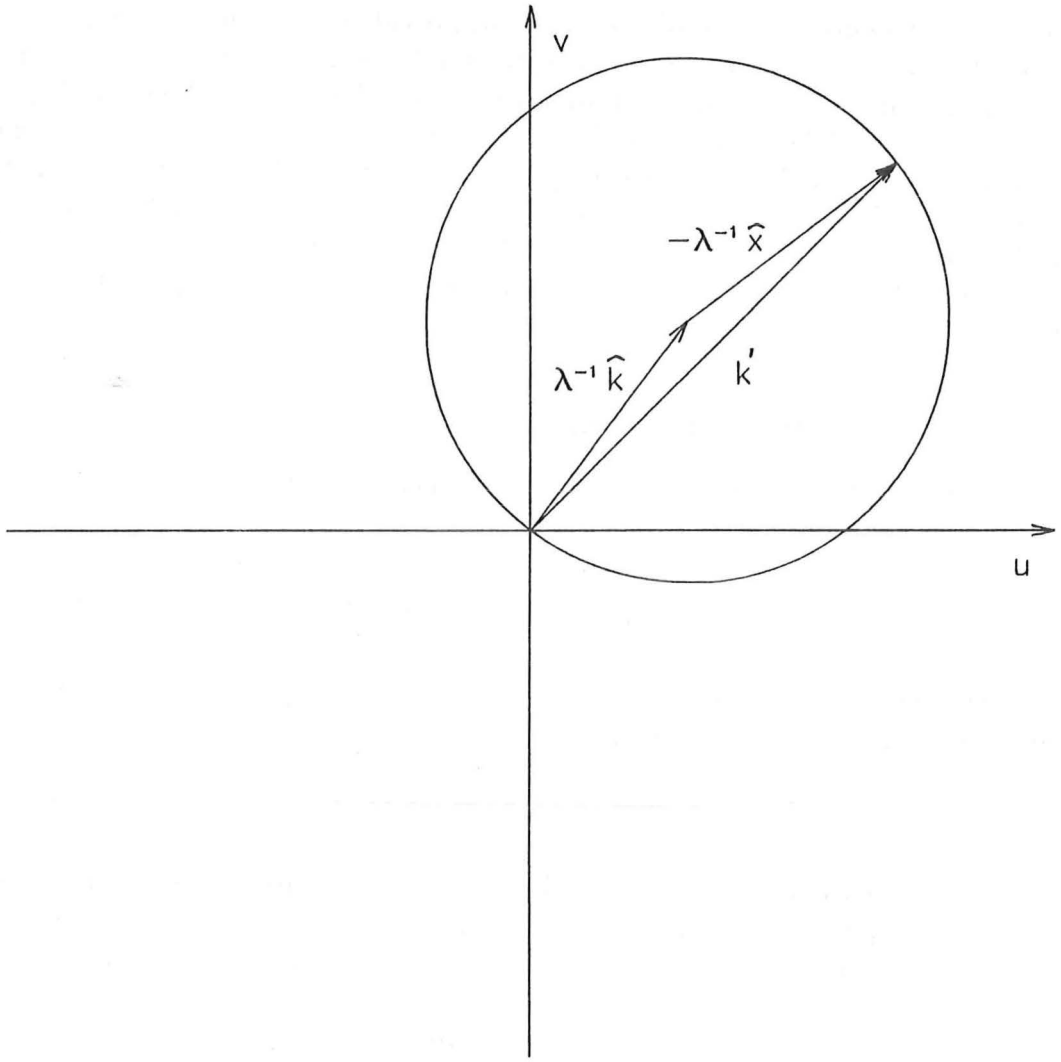

Figure 4.1: Fourier transform relation within the Born approximation. The far scattered field is related to the portions of the Fourier transform of χ on a circle in Fourier space. The centre of the circle is determined by the direction $\vec{k}$ of the incident field.

In the event that the incident field is a plane-wave with direction of propagation $\vec{k}$, with $|\vec{k}| = k$, so that

$$\psi_i(\vec{x}, k) = e^{-i\vec{k}\cdot\vec{x}}, \tag{4.9}$$

it is seen that the approximation to the scattered field given by the first term in the Neumann series, introduced in Section 4.2 above, which is called the first Born approximation (also known as the Rayleigh–Gans approximation [Jones 1964, Chapter 6]), becomes (by setting $N = 1$ in (4.7))

$$\psi^{(B,1)}(\vec{x}, k) = \int_{\Upsilon_s} \left[\chi^2(\vec{x}_1, k) - 1\right] e^{-i(\vec{k}-k\hat{x})\cdot\vec{x}_1} d\Upsilon_s(\vec{x}_1). \tag{4.10}$$

On defining

$$\vec{k}' = \lambda^{-1}(\hat{k} - \hat{x}), \tag{4.11}$$

and using the notation of Section 2.7, the (first) Born approximation in the far-field can be written as

$$\psi^{(B,1)}(\vec{x}, k) = \mathcal{F}[\chi^2 - 1](\vec{k}'). \tag{4.12}$$

That is, the value of $\psi^{(B,1)}$ around a circle in the far-field is proportional to the Fourier transform of $(\chi^2 - 1)$ on an offset circle in Fourier space of radius λ^{-1} (see Figure 4.1) [Wolf 1969].

The Neumann series is known as the Born series when the conventional Green's function is employed. Higher order Born approximations are constructed by taking account of increasing numbers of terms in the Neumann series.

The condition (4.6) is sufficient to ensure convergence of the Born series and is often readily evaluable, but is unnecessarily restrictive. It is, however, an appropriate criterion for gauging whether the first Born approximation is a useful approximation to the scattered field. For, not only must the incident field be much greater in amplitude than the scattered field, but it must suffer negligible phase delay (no more than $\pm\pi/2$ radians, say) on its passage through the inhomogeneity [Slaney *et al.* 1984]. Note that the assumptions of the echo-location model are precisely that the Born approximation is valid.

Despite a lack of justification in many instances, the Born approximation remains one of the most commonly invoked approximations [Blackledge *et al.* 1987a], [Blackledge *et al.* 1987b], [Boyse and Keller 1986], [Norton 1983], due to the ease with which its inverse (discussed in Section 6.5) can be formed.

4.2.2 Rytov approximations

The Rytov approximation is obtained by a similar approach to that of the previous section. First, new variables are introduced [Ishimaru 1978, Chapter 17] :

$$\psi = \exp(\gamma), \tag{4.13a}$$

$$\gamma = \gamma_i + \gamma_s, \tag{4.13b}$$

$$\gamma_i = \ln \psi_i. \tag{4.13c}$$

Elementary calculus reveals that

$$\nabla\psi = \psi\nabla\gamma, \tag{4.14}$$

$$\nabla^2\psi = \psi(\nabla^2\gamma + \nabla\gamma \cdot \nabla\gamma). \tag{4.15}$$

Furthermore, the free-space Green's function is employed, and ψ_i is taken to to be a plane-wave of the form (4.9), implying that Σ can be taken to be zero. Substitution of (4.15) into (2.78) yields

$$\nabla^2\gamma + \nabla\gamma \cdot \nabla\gamma + k^2\chi^2 = 0. \tag{4.16}$$

But (4.13b) and (4.13c) imply that

$$\begin{aligned} \nabla\gamma &= \nabla \ln\psi_i + \nabla\gamma_s \\ &= \frac{\nabla\psi_i}{\psi_i} + \nabla\gamma_s, \end{aligned} \tag{4.17}$$

$$\begin{aligned} \nabla^2\gamma &= \frac{\nabla^2\psi_i}{\psi_i} - \frac{\nabla\psi_i \cdot \nabla\psi_i}{\psi_i^2} + \nabla^2\gamma_s \\ &= -k^2 - \frac{\nabla\psi_i \cdot \nabla\psi_i}{\psi_i^2} + \nabla^2\gamma_s, \end{aligned} \tag{4.18}$$

where (3.2) has been applied to ψ_i. Thus, (4.16) becomes

$$\nabla^2\gamma_s + 2\nabla\gamma_i \cdot \nabla\gamma_s = -\left[k^2(\chi^2 - 1) + \nabla\gamma_s \cdot \nabla\gamma_s\right]. \tag{4.19}$$

Now put

$$\phi = \gamma_s\psi_i, \tag{4.20}$$

$$\nabla\phi = \psi_i\nabla\gamma_s + \gamma_s\nabla\psi_i, \tag{4.21}$$

$$\begin{aligned} \nabla^2\phi &= \psi_i\nabla^2\gamma_s + 2\nabla\gamma_i \cdot \nabla\gamma_s + \gamma_s\nabla^2\psi_i \\ &= \psi_i\left(\nabla^2\gamma_s + \frac{2}{\psi_i}\nabla\psi_i \cdot \nabla\gamma_s\right) - k^2\psi_i\gamma_s \\ &= \psi_i(\nabla^2\gamma_s + 2\nabla \ln\psi_i \cdot \nabla\gamma_s) - k^2\phi, \end{aligned} \tag{4.22}$$

so that (4.19) becomes

$$\nabla^2\phi + k^2\phi = -\left[k^2(\chi^2 - 1) + \nabla(\frac{\phi}{\psi_i}) \cdot \nabla(\frac{\phi}{\psi_i})\right]\psi_i. \tag{4.23}$$

The linear equation (2.78) has thus been transformed to a nonlinear differential equation involving the product of two gradient terms.

The conventional volume-source formulation (see Section 3.4) for (4.23) involves a nonlinear integral operator

$$K^{(R)}(\phi) = K^{(B)}\psi_i + \int_{\Upsilon_s} \nabla(\frac{\phi}{\psi_i}) \cdot \nabla(\frac{\phi}{\psi_i})\psi_i g \, d\Upsilon_s, \tag{4.24}$$

where the operator $K^{(B)}$ is defined by (4.5), converting (4.23) into a nonlinear integral equation

$$\phi = K^{(R)}(\phi), \tag{4.25}$$

for ϕ. The Rytov approximation is obtained by inserting the initial estimate

$$\phi^{(R,0)} = 0, \tag{4.26}$$

into (4.25), resulting in an expression of exactly the same functional form as the Born approximation :

$$\begin{aligned} \phi^{(R,1)} &= K^{(R)}\left(\phi^{(R,0)}\right) \\ &= K^{(B)}\psi_i. \end{aligned} \tag{4.27}$$

The Born and Rytov approximations are distinguished by their different interpretations of the quantity $K^{(B)}\psi_i$ which, in the Born approximation, represents an approximation to the scattered field ψ_s. In the Rytov approximation, however, the expression represents the quantity ϕ which is related to the total field through

$$\phi = \psi_i \ln(\psi/\psi_i). \tag{4.28}$$

Defining higher order approximations recursively,

$$\phi^{(R,N)} = K^{(R)}\left(\phi^{(R,N-1)}\right), \tag{4.29}$$

is reminiscent of the fixed-point algorithm to solve nonlinear equations [Rall 1969]. Weston [1985] exploits this to derive conditions under which the Rytov approximation becomes the leading term in a series converging to the exact solution of (4.25). Keller [1969] interprets the Rytov approximation as an asymptotic expansion, superior to the Born approximation in the forward direction, but inferior when reflection is significant.

A class of transformations has been proposed [Kaveh *et al.* 1982] incorporating both the Born and Rytov approximations as special cases. Bates *et al.* [1976] describe expressions which partially compensate for the term $\nabla\left(\frac{\phi}{\psi_i}\right)\cdot\nabla\left(\frac{\phi}{\psi_i}\right)$ in (4.23).

The far-field form of the (first) Rytov approximation for the complex phase function is identical to the expression obtained for the far-field form of the (first) Born approximation, i.e.

$$\phi^{(R,1)}(\vec{x},k) = \mathcal{F}[\chi^2 - 1](\vec{k}'), \tag{4.30}$$

using the notation of Section 4.2.1.

4.3 NULL-FIELD METHOD

The null-field method [Bates 1968,Bates and Wall 1977a,Varadan and Varadan 1980], which is an adaptation of Waterman's [1965,1969] extended boundary condition (known in optical contexts as the Ewald-Oseen extinction theorem [Lalor and Wolf 1972], [Pattanayak and Wolf 1972]), is a formulation of the direct problem that is based on the surface-source description of scattering introduced in Section 3.4.2. The null-field method has been successfully applied to a multitude of direct scattering problems in electromagnetic, acoustic [Peterson and Ström 1975], and elastic wave motion [Pao and Varatharajulu 1976,Varatharajulu and Pao 1976,Varadan *et al.* 1978]. The null-field method appropriate for problems involving an impenetrable object governed by the Dirichlet boundary condition, first introduced in Section 2.5.1, is now formulated in terms of any one of the 11 coordinate systems in which the Helmholtz equation separates. The null-field method best suited numerically for a particular scattering obstacle is the one that employs the coordinate system in which the boundary of the

obstacle most closely corresponds to a coordinate surface [Lewin 1970]. In preparation for the inverse problem considered in Appendix 6-B, the general null-field equations are stated explicitly for the specialisation to cylindrical polar coordinates under the further assumption that the direct problem is independent of z, as is appropriate for two-dimensional problems. Modifications of the procedure derived here to incorporate the other class of impenetrable objects, i.e. those for which the Neumann boundary condition applies, involve only minor variations in the algebra (see, for example, Bates and Wall [1977a]).

Consider, then, an impenetrable obstacle embedded in an otherwise homogeneous medium whose boundary, denoted S, is a simply connected closed curve that partitions Υ into an interior region Υ_s, containing the origin O, and an exterior region Υ_+ which is assumed to be homogeneous with wavespeed normalised to unity. It is assumed that S is star-like, i.e. an arbitrary point on S can be represented by

$$\xi_1' = \xi_1'(\xi_2', \xi_3'), \tag{4.31}$$

where the notation of Section 3.3.1 applies throughout this section.

The wave motion is governed by the equations of Section 3.4.2 which, when the Dirichlet boundary condition applies, simplify to

$$\oint_S g(|\vec{x} - \vec{x}'|, k) \frac{\partial u}{\partial n}(\vec{x}', k)\, dS(\vec{x}') = \begin{cases} -\psi_i(\vec{x}, k), & \vec{x} \in \Upsilon_s, \\ \psi_s(\vec{x}, k), & \vec{x} \in \Upsilon_+. \end{cases} \tag{4.32}$$

It suffices to assume that the incident field is regular throughout all of Υ, so that its angular variation is appropriately expressed in the form (3.30), while its radial variation is also comprised of regular functions,

$$\psi_i(\vec{x}, k) = \sum_m a_m y_m^{(r)}(\xi_1) Y_m(\xi_2, \xi_3). \tag{4.33}$$

For $\vec{x} \in \Upsilon_s$, (3.32) applies with $\xi_{1>} = \xi_1'$ and $\xi_{1<} = \xi_1$. The right hand side of (4.32) is in this case known to the experimenter, making (4.32) an integral equation for an equivalent surface source density. Invoking (4.31) to parametrise S, and recalling (3.32), an infinite set of moment equations can be obtained by applying the "test" functions $\{C_m Y_m^*\}$, and the Petrov-Galerkin method [Harrington 1968, Chapter 1]. The set $\{C_m\}$ are constants chosen to improve the numerical conditioning of the resulting null-field equations, so obtained,

$$C_m \iint f(\xi_2', \xi_3') c_m(\xi_2', \xi_3') y_m^{(o)}(\xi_1'(\xi_2', \xi_3')) Y_m^*(\xi_2', \xi_3')\, d\xi_2'\, d\xi_3' = C_m a_m, \quad \text{for all } m, \tag{4.34}$$

where the surface source density f depends on S:

$$f(\xi_2', \xi_3') = \frac{\partial \psi}{\partial n}(\xi_1'(\xi_2', \xi_3'), \xi_2', \xi_3', k). \tag{4.35}$$

Adopting the expansion (3.30) for f,

$$f(\xi_2', \xi_3') = \sum_m f_m Y_m(\xi_2', \xi_3'), \tag{4.36}$$

the null-field equations reduce to solving the infinite system of linear equations

$$\sum_n H_{mn} f_n = -U_m, \quad \text{for all } m, \tag{4.37}$$

where, for each pair m, n,

$$H_{mn} = C_m \iint y_m^{(o)}(\xi_1'(\xi_2', \xi_3'))c_m(\xi_2', \xi_3')Y_n(\xi_2', \xi_3')Y_m^*(\xi_2', \xi_3')\, d\xi_2'\, d\xi_3'. \tag{4.38}$$

When $\vec{x} \in \Upsilon_s$, $\xi_{1>} = \xi_1$ and $\xi_{1<} = \xi_1'$, so that the arguments of $y_m^{(r)}$ and $y_m^{(o)}$ occurring in (3.32) as employed in (4.32) are interchanged. The scattered field then has the form

$$\psi_s(\xi_1, \xi_2, \xi_3, k) = \sum_m b_m y_m^{(o)}(\xi_1)Y_m(\xi_2, \xi_3), \tag{4.39}$$

where the b_m are given by

$$b_m = \sum_n J_{mn} f_n, \tag{4.40}$$

and

$$J_{mn} = \iint y_m^{(r)}(\xi_1'(\xi_2', \xi_3'))c_m(\xi_2', \xi_3')Y_n(\xi_2', \xi_3')Y_m^*(\xi_2', \xi_3')\, d\xi_2'\, d\xi_3'. \tag{4.41}$$

As a special case, consider the development for two-dimensional fields in cylindrical coordinates. The radial distance from the origin to a point on the boundary of the obstacle is conveniently expressed as a function of angle, i.e.

$$\rho' = \rho'(\phi'). \tag{4.42}$$

The angular harmonics are given by (3.33), so that $c_m = 1$ for all m, and

$$f(\phi') = \sqrt{\rho'^2(\phi') + \left(\frac{d\rho'}{d\phi'}\right)^2}\, \frac{\partial u}{\partial n}(\rho'(\phi'); \phi', k). \tag{4.43}$$

Since f has period 2π, it is appropriately expanded according to (4.36),

$$f(\phi') = \sum_{m=-\infty}^{\infty} f_m e^{im\phi'}. \tag{4.44}$$

In the far-field, the scattered field reduces to the asymptotic form

$$\psi_s \sim 2(\pi kr)^{-1/2} \exp(-ikr) \sum_{m=-\infty}^{\infty} b_m i^m e^{im\phi}, \tag{4.45}$$

and, consequently,

$$H_{mn} = C_m \int_0^{2\pi} H_m^{(2)}(k\rho'(\phi')) \exp(i(n-m)\phi')\, d\phi', \tag{4.46}$$

$$J_{mn} = \int_0^{2\pi} J_m(k\rho'(\phi')) \exp(i(n-m)\phi')\, d\phi'. \tag{4.47}$$

Note that, since J_m is the real part of $H_m^{(2)}$, much of the computation involved in evaluation of integrals in (4.46) and (4.47) is the same.

When solving the direct problem for a particular $\rho'(\phi')$, the conditioning of (4.37) is improved by suitable normalisation of the Hankel functions appearing in (4.38) [Wall 1980]. Equivalently, this may be accomplished by choosing the constants

C_m to be $1/H_m^{(2)}(k\rho'(0))$. Thus, in the event that S is a circle of radius ρ_0', H_{mn} reduces to the Kronecker delta, and [Wall *et al.* 1985]

$$f_m = (-i)^m / H_m^{(2)}(k\rho_0'), \tag{4.48}$$

$$b_m = (-i)^m J_m(k\rho_0') / H_m^{(2)}(k\rho_0'), \tag{4.49}$$

as can be deduced from separation of variables [Bowman *et al.* 1969].

In practice, the Fourier series are truncated to a finite number of terms, as are the matrices H and J.

4.4 PHYSICAL OPTICS

The major computational effort associated with the null-field method described in the previous section is the solution of an integral equation to determine the equivalent source density f, existing at points Q on the surface S of the scatterer, which radiates the same scattered field as the actual scatterer exterior to Υ_s. The method is, in principle, exact. Considerably less effort is required when an approximate equivalent source $\Sigma_s(Q)$ is employed, whose value at each point Q can be postulated on the basis of an intuitive assumption, which is thought (because of the physics of whatever situation is being considered) likely to lead to a useful degree of approximation. In the physical optics approximation, the equivalent source is taken to be that which is exact for a body, here called the assumed body, belonging to a particular class of bodies of certain specified constitution. Furthermore, the tangent planes at Q of the actual body and the assumed body are required to coincide [Bates and Wall 1977b]. This assumption may seem quite unwarranted at first sight. However, it often leads to results in remarkable agreement with experiment.

There are several species of physical optics. The best known, most commonly referred to as merely physical optics [Senior 1965] or the Kirchoff approximation [Bouwkamp 1954], is what Bates and Wall [1977b] call planar physical optics. The assumed body has an infinite planar surface, and is usually totally reflecting [Bleistein and Cohen 1980], but can in fact have any homogeneous constitution since exact expressions can be derived for such scatterers [Jones 1964, Chapter 6]. The equivalent source assumes non-zero values only in the region directly illuminated by the incident wave, while within the penumbra and umbra it is set to zero (see Figure 4.2a).

Planar physical optics can give highly accurate results for actual bodies of large transverse extent such as the ocean surface or a large aircraft viewed from above or below. It tends to be much less satisfactory when scattering from the penumbra of the actual body becomes significant, and for objects whose largest transverse dimensions are a few wavelengths or less. One can then employ other forms of physical optics [Bates and Wall 1977b] in which the assumed body surface coincides with more general surfaces defined by coordinate systems in which the Helmholtz equation separates (refer to Section 3.3.1). In such a coordinate system, it is necessary to generalise the notion of directly illuminated and shadow regions. The directly illuminated region consists of all points $(\xi_1^{(d)}, \xi_2, \xi_3)$ on S such that, for particular values of ξ_2 and ξ_3, (ξ_1, ξ_2, ξ_3) is exterior to Υ_s whenever $\xi_1 > \xi_1^{(d)}$. The shadow region consists of all other points of S (see Figure 4.2b).

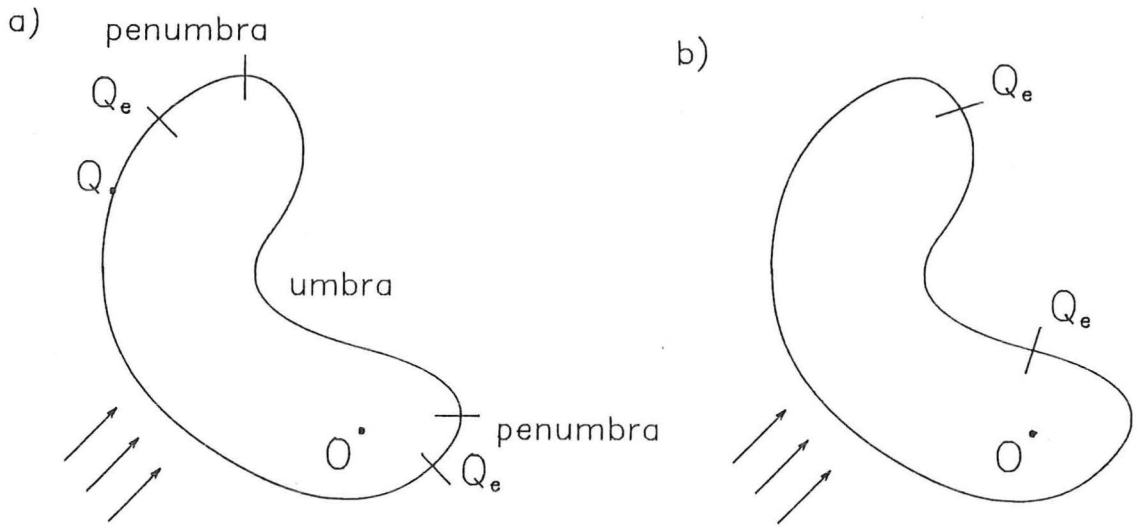

Figure 4.2: Physical optics geometry. A plane wave is incident on a scatterer containing the origin O. (a) In planar physical optics, the equivalent source density is zero throughout the umbra and penumbra. The illuminated region contains points Q and extends to Q_e. (b) Other versions of physical optics permit more faithful approximation source densities over a generalised illuminated region, whose extent is increased.

Chapter 5

PRE-PROCESSING FOR INVERSION ALGORITHMS

5.1 INTRODUCTION

This chapter and the next are concerned with the third task involved in examining an inverse problem (refer to Section 1.4). Relationships between emanations and the generalised constitutive parameter, established in previous chapters in the context of the direct problems, are reconsidered from the point of view of inverse problems. The particular concern is with devising algorithms to infer the generalised constitutive parameter from whatever functional of the scattered field represents the data available in an inverse problem. Since the great majority of solutions to inverse problems involve intricate computations, a critical consideration is the numerical stability of any algorithm which infers the generalised constitutive parameter from given values of its scattered field. It must be remembered that inverse problems are often ill-posed (refer to Section 1.5). Consequently, some form of regularisation is required, either explicitly (i.e. replacing the original problem by a well-posed one) or implicitly (i.e. by considering a suitable finite dimensional space to represent the unknowns).

Mathematically, the relations of Chapter 4 are viewed notionally as examples of the mapping (4.1) from a space of generalised constitutive parameters and incident fields, to a space characterising the scattered fields. For the corresponding inverse problem, the roles of the generalised constitutive parameter and the scattered field are reversed. In both problems, the incident field is under the control of the experimenter (refer to Section 1.3). The process of solving an inverse problem can thus be written in terms of evaluating the inverse mapping

$$\Omega^{-1} : E \to C. \tag{5.1}$$

A thorough understanding of the well- or ill- posedness of an inverse problem depends on the nature of the inverse map, which in turn depends on the characterisation of the spaces E and C. Note that an ill-posed problem can often be converted to a well-posed problem by the incorporation of *a priori* information that serves to regularise the original problem. The Newton-Kantorovich method (see Section 6.7) is particularly amenable to such regularisation.

It only makes sense to ask whether an inverse problem is ill-posed if the physical model underlying the associated direct problem and its mathematical description are

sufficiently simple and well understood that the components of the inverse problem, i.e. the spaces C and E, and also Ω, can be characterised to the extent that it is meaningful to ask the question in the first place. There are a large number of applications in which the question seems inappropriate due to lack of adequate knowledge of any one of these components.

Useful solutions to inverse problems are only possible in practice if satisfactory data sets are available. The techniques presented in Chapter 6 are predicated on the assumption that all required data are available. In the real world, however, data sets are usually incomplete. Often, the quantity of data is sufficient, but the quality is inadequate in some sense. If the data contamination is due to random noise, of a level which is excessive for the purposes in hand, then nothing can of course be done. There are many situations, however, wherein the deficiencies of the data happen to be such that they can be rectified to a useful degree by either appropriate preprocessing or invoking *a priori* information or both. The remainder of this chapter discusses situations in which the data need considerable preliminary massaging before they can be operated on usefully by an inversion algorithm. The particular instances examined concern data that are :

(a) explicitly complete in quantity but blurred (Section 5.2).

(b) apparently incomplete, in the sense that complete data sets are not immediately available, but can (sometimes) be inferred (Sections 5.3 and 5.4).

(c) palpably incomplete, wherein the data cannot be completed in any reliable manner, although the use of heuristics often leads to results that can be interpreted by trained practitioners (Section 5.4).

Section 5.2 is concerned with the archetypal problem of type (a), namely, deconvolution.

One way that the data sets of type (b) can be made satisfactory is by invoking *a priori* information which allows the invocation of theoretical constraints. This is demonstrated in Section 5.3 in the context of retrieving phases of scattered fields from magnitudes. This problem area is one of the most active presently under investigation by the image processing community [Bates and Mnyama 1986], [Fienup and Wackerman 1986]. Since the majority of data sets for inverse problems are obtained from measurements on wave fields, the phases of data are as important as their intensities. As is explained in Section 5.3, there are certain severe practical impediments in the way of retrieving inverse scattering phases. The published paper [Bates and Tan 1985a] reproduced as Appendix 5-A was perhaps the earliest to emphasise the computational niceties of these impediments. The published paper [Bates and Tan 1985b] reproduced as Appendix 5-B was the first to show how to invoke extra information (which can be expected to be readily available in practice) that allows the impediments to be circumvented with some ease. The published paper [Gardenier *et al.* 1986] reproduced as Appendix 5-C demonstrates another way in which *a priori* information, despite being possibly inexact, can be incorporated into a phase retrieval algorithm. The published paper [Bates and Tan 1986] reproduced as Appendix 5-D applies phase retrieval techniques to a relatively new discovery in material science [Shechtman *et al.* 1984] that has aroused considerable debate [Pauling 1985].

Section 5.4 is concerned with the models employed in a class of imaging techniques called computed tomography (CT). The archetypal CT model is predicated upon a description of the emanations as travelling along straight rays. Although the projection theorem is the basis of the remarkably successful reconstruction algorithms discussed in Section 6.3, it is convenient to introduce it in this chapter. The projection model has achieved remarkable success when applied to imaging the X-ray attenuation coefficient of opaque materials, for the purposes of medical diagnostic radiology. Section 5.4 discusses methods for approximately transforming projection data so that they correspond to straight-line projections, even though the data are not inherently of such form. Instances of type (b) and (c) are included. The particular example of incompleteness examined in Appendix 5-E concerns the effects on reconstructions of detector resolution and finite sampling of projections.

To accord with the terminology of this thesis, the quantity of interest $f(\vec{x})$ is referred to throughout this Chapter as the image [Bates and Mnyama 1986], although the term "object" is also popular in phase retrieval contexts [Fienup 1983b], [Arsenault and Chalasinska-Macukow 1983], [Fienup 1984], [Byrne and Fiddy 1987].

5.2 DECONVOLUTION

One of the ways in which practical reality falls short of mathematical ideality is apparent when one remembers that data are acquired by some transduction process (refer to Section 1.2). Rather than consisting of precise measurements, available data sets are often "blurred" by "point spread functions" or "instrument functions" [Goodman 1968, Chapter 6], [Bracewell 1978, Chapter 3], [Bates and McDonnell 1986, Sections 3–4]. Although such blurring is itself ideal, because it is point spread invariant or isoplanatic, it occurs to a good approximation at least in many situations of scientific and technological importance. Before attempting to operate on such data with inversion algorithms, it is appropriate to preprocess them to remove as much of the blurring as possible.

An appropriate description of the relation between isoplanatically blurred data $s(\vec{x})$ and the data of interest $f(\vec{x})$ is supplied by a convolutional model

$$s(\vec{x}) = f(\vec{x}) \odot h(\vec{x}) + c(\vec{x}), \tag{5.2}$$

where $h(\vec{x})$ is known as the point-spread, or blurring, function, and $c(\vec{x})$ is the contamination. The Fourier transforms of s and f are related through

$$S(\vec{u}) = F(\vec{u})H(\vec{u}) + C(\vec{u}), \tag{5.3}$$

which follows from the formulas presented in Section 2.7.

In those situations wherein the blurring is negligible, the point-spread function effectively reduces to $\delta(\vec{x})$, or equivalently $H(\vec{u})$ becomes unity. When, as is perhaps more usual, the blurring is significant, it is necessary to estimate the form of $h(\vec{x})$ before one can recover $f(\vec{x})$. Current research [Lane and Bates 1987a], [Lane *et al.* 1987], [Lane and Bates 1987b] suggests that it may soon always be possible to do without prior knowledge of $h(\vec{x})$. At present, however, when presented with a single blurred image $s(\vec{x})$, a prior estimate of $h(\vec{x})$ is in general needed. From this, one can form the

"Wiener filter" [Bates and McDonnell 1986, Section 16]

$$W(\vec{u}) = H^*(\vec{u})/[|H(\vec{u})|^2 + \Phi(\vec{u})], \tag{5.4}$$

where $\Phi(\vec{u})$ characterises the available knowledge of the dependence upon $\vec{u}$ of the noise-to-signal ratio for the data. An estimate $\hat{F}(\vec{u})$ of $F(\vec{u})$ is then generated :

$$\hat{F}(\vec{u}) = F(\vec{u})W(\vec{u}). \tag{5.5}$$

Note that, in the absence of contamination, $W(\vec{u})$ reduces to the "inverse filter" $1/H(\vec{u})$, implying that $\hat{F}(\vec{u})$ becomes identical to $F(\vec{u})$. It is also worth remembering that $\Phi(\vec{u})$ can incorporate differences between the actual $h(\vec{x})$ and whatever estimate of it is available.

Wiener filtering can be performed when only a single blurred image is available, provided a reasonable estimate of H is available. The dominant component of (5.2) must be the convolution, with the contamination much less significant. Often, the contamination is so severe that deconvolution is impracticable unless a number of differently blurred images are available. There is a class of image processing techniques, originally devised in astronomical imaging contexts [Labeyrie 1970], [Bates 1982], that can be usefully invoked on occasions in which an ensemble of blurred images is given, rather than to a single convolution. The blurring has a "speckly" character due to the way in which the images are formed [Dainty 1974], [Dainty 1975b], [Goodman 1975]. The speckle images, as they are commonly called, are assumed to be due to point-spread functions that are statistically independent. Prior estimates of these point spread functions are not required. The m-th speckle image is expressed as a convolution of the true image with a point-spread function $h_m(\vec{x})$ plus a contamination term $c_m(\vec{x})$, i.e.

$$s_m(\vec{x}) = f(\vec{x}) \odot h_m(\vec{x}) + c_m(\vec{x}). \tag{5.6}$$

Denote the point at which s_m has greatest magnitude by $\vec{x}_m$. The technique known as shift-and add (SAA) forms an image f_{SAA} by shifting each speckle image so that its brightest point is positioned at the origin of image space, normalising each so that the phase at the origin is zero, and then averaging the shifted images [Bates and Cady 1980], [Bates and Minard 1984]. Denoting this reconstruction by f_{SAA}, it is seen that

$$\begin{aligned} f_{\mathrm{SAA}}(\vec{x}) &= \langle s_m(\vec{x} + \vec{x}_m) s_m^*(\vec{x}_m)/|s_m(\vec{x}_m)| \rangle \\ &= f(\vec{x}) \odot \langle h_m(\vec{x} + \vec{x}_m) \rangle + \langle c_m(\vec{x} + \vec{x}_m) \rangle \,, \end{aligned} \tag{5.7}$$

where $\langle \cdot \rangle$ denotes the ensemble average.

The usefulness of SAA in any particular application depends primarily on the extent to which the convolutional component of (5.6) dominates the contamination term. It is now well appreciated that the technique can be very effective for imaging spatially incoherent source distributions [Bates 1982], [Baba *et al.* 1985], [Freeman *et al.* 1988], [Bates and Davey 1988]. It is becoming apparent that SAA can also be usefully invoked as a post-processing procedure to yield faithful images of objects which emit quasi-monochromatic fields exhibiting a high degree of spatial coherence [Minard *et al.* 1985]. This is discussed further in Section 6.6.

5.3 PHASE RETRIEVAL

It is implicit in the inversion algorithms considered in the next chapter that the data available are in the form of measurements of both the magnitude and phase of the scattered field. There are a number of applications in which measurement of the phase of the scattered field can only be performed with more difficulty (expense), or to much less accuracy, than measurement of magnitude [Gerchberg and Saxton 1971], [Fright 1984, Chapter 2], [Morris 1985], [Anderson and Sali 1985]. Measurement of the phase can even be impractical [Ramachandran and Srinivasan 1970, Chapter 1], [Misell 1973], [Ferwerda 1978], [Bates 1982]. In such cases, it is useful to attempt to infer the phase from the magnitude of the scattered field.

It is often the case that the quantity available is the magnitude $|F(\vec{u})|$ of the Fourier transform of $f(\vec{x})$. The problem of inferring the phase of F is then referred to as Fourier phase retrieval [Bates and McDonnell 1986, Chapter 4].

It is instructive to define an equivalence class [Herstein 1975, Chapter 1] of an image $f(\vec{x})$, containing all translations $f(\vec{x}-\vec{\alpha})$, any multiplication by a complex constant $\beta f(\vec{x})$, and all shifted conjugate reflections $f^*(\vec{\alpha}-\vec{x})$. All members of an equivalence class possess the same Fourier transform modulus, apart from an overall multiplicative factor, and hence a phase retrieval algorithm cannot distinguish between them. The absolute position of the image and its orientation are irretrievably lost, but they can often be inferred by other means and are equally often relatively unimportant [Bates and Fright 1983,Bates and Mnyama 1986] This non-uniqueness is usually disregarded because all images belonging to an equivalence class have the same appearance, or "image-form" [Bates and Fright 1983]. The phase retrieval problem is more meaningfully posed as : recover the image-form of f from $|F|$.

The support of f, denoted $supp(f)$, is defined as the set of points, $\vec{x}$, at which $f(\vec{x})$ is non-zero (see Section 2.7). It is assumed that the support of f is bounded, which is certainly valid for all of the applications envisioned in this section. Denoting the maximum linear extent of $supp(f)$ in each Cartesian direction by L_x and L_y respectively, the sampling theorem (refer to Section 2.7) states that f is determined by samples of F (of both magnitude and phase) spaced by $1/L_x$ and $1/L_y$ in each of the corresponding Fourier space variables, u and v. The autocorrelation ff of f has extent no greater than $2L_x$ and $2L_y$, and is determined by samples of its Fourier transform $|F|^2$ (see (2.119)), spaced by no more than $1/2L_x$ and $1/2L_y$. It is assumed in this section that $|F|$ is sampled by at least the latter rate so that ff can be inferred immediately from the data, and an estimate of $supp(f)$ obtained by halving the linear dimensions of $supp(ff)$.

The Fourier phase problem is, equivalently, posed as finding the image-form of f from its autocorrelation ff. The issue of uniqueness of the image-form for an image composed of δ-functions weighted by complex numbers,

$$f(\vec{x}) = \sum_{m_1=-M_1}^{M_1} \cdots \sum_{m_K=-M_K}^{M_K} f_{m_1 \ldots m_K} \delta(x_1 - m_1, \ldots, x_K - m_K), \tag{5.8}$$

can be most readily appreciated by introduction of the complex $\vec{z}$-transform

$$P(\vec{z}) = \sum_{m_1=-M_1}^{M_1} \cdots \sum_{m_K=-M_K}^{M_K} f_{m_1 \ldots m_K} {z_1}^{m_1} \ldots {z_K}^{m_K}. \tag{5.9}$$

Each component of $\vec{z}$ is a complex number. The $\vec{z}$-transform and Fourier transform of f are related by

$$\mathcal{F}[f](\vec{u}) = P(e^{-i2\pi\vec{u}}), \tag{5.10}$$

where the exponentiation is to be understood component-wise. The image-form of f is completely characterised by the surface in complex $\vec{z}$-space on which $P(\vec{z}) = 0$. This surface, which is of dimension $(2K-2)$, and is embedded in a $2K$-dimensional complex $\vec{z}$-space (i.e. 2 real dimensions for each complex component of $\vec{z}$), is known as the zero-sheet of $P(\vec{z})$ [Lane and Bates 1987a,Lane *et al.* 1987]. The $\vec{z}$-transform of ff is, from (2.118),

$$PP(\vec{z}) = P^*(\tilde{z})P(\vec{z}), \tag{5.11}$$

where the components of $\tilde{z}$ are $\tilde{z}_l = 1/\bar{z}_l$, and $\bar{z}_l$ is the ordinary complex conjugate of z_l. On denoting the zero-sheet of $P(\vec{z})$ by Z_P, the zero-sheet of $PP(\vec{z})$ is given by

$$Z_{PP} = Z_P \cup Z_{P^*}. \tag{5.12}$$

In one dimension, it is well known that even the image-form is not unique [Bruck and Sodin 1979]. The zero-sheets are merely 0-dimensional, i.e. a set of discrete points. The z-transform of f reduces to an ordinary polynomial which can always be factored,

$$P(z) = \sum_{m=-M}^{M} f_m z^m = Z \prod_{m=0}^{2M} (z - z_m), \tag{5.13}$$

where Z is an unimportant term proportional to z^{-M}. The z-transform of ff is, likewise,

$$PP(z) = P^*(\tilde{z})P(z) = |Z|^2 \prod_{m=0}^{2M} (z - z_m)(z - 1/\bar{z}_m). \tag{5.14}$$

The zeros of PP can be inferred from ff, or equivalently $|F|^2$ (because of analytic continuation), and form a discrete set of pairs in the complex z-plane. There is clearly no way to deduce which of each pair belong to f and which to its mirror image. There are instances where a number of selections are still consistent with other constraints such as positivity (see Section 2.7) of f.

The situation in more than one dimension is completely different. Each zero-sheet forms a continuous (in fact, analytic) sheet. The intersections of Z_P and Z_{P^*} in Z_{PP} occur at points. However, if one imagines a continuous line passing from Z_P to Z_{P^*} through any such point, contained entirely within Z_{PP}, this line ceases to be analytic at the said point. This makes it possible to construct an algorithm for separating Z_P from Z_{P^*}, and hence to infer $f(\vec{x})$ [Lane *et al.* 1987]. The computational and practical implications have yet to be fully appreciated. Consequently, it is still necessary to invoke iterative procedures, of the kind discussed below, in order to effect practical phase retrieval. However, the concept of the zero-sheet graphically demonstrates when the multi-dimensional phase retrieval can in general be expected to possess a unique solution (image-form).

The main concern of this thesis is multi-dimensional inverse scattering. The remainder of this section considers problems that involve more than one dimension.

The techniques for phase retrieval that are simplest to invoke are iterative algorithms [Gerchberg and Saxton 1972], [Fienup 1978], [Fienup 1979], [Fienup 1980],

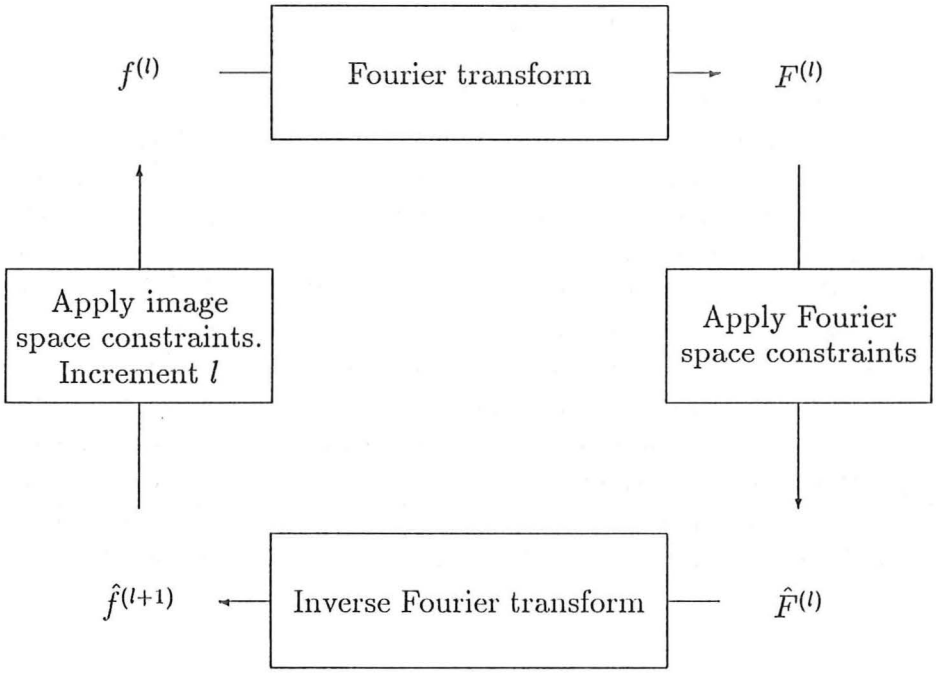

Figure 5.1: A typical Fienup iteration. The cycle proceeds clockwise. Horizontal arrows represent (inverse) Fourier transformations. Vertical arrows represent imposition of other constraints.

[Fienup 1982], in which it is relatively easy to incorporate other information. The essential components of a Fienup iteration are shown schematically in Figure 5.1. An estimate $f^{(l)}$ of the image is Fourier transformed to produce $F^{(l)}$. Fourier space constraints are then applied to $F^{(l)}$ to yield $\hat{F}^{(l)}$. An inverse Fourier transform results in $\hat{f}^{(l+1)}$. Finally, image-space constraints are imposed upon $\hat{f}^{(l+1)}$ to generate $f^{(l+1)}$ which completes one iteration. The respective relationships between $f^{(l)}$ and $F^{(l)}$, and $\hat{F}^{(l)}$ and $\hat{f}^{(l+1)}$, are

$$F^{(l)} = \mathcal{F}[f^{(l)}], \tag{5.15}$$

$$\hat{f}^{(l+1)} = \mathcal{F}^{-1}[\hat{F}^{(l)}]. \tag{5.16}$$

The Fourier space constraint most commonly applied is the combination of the known Fourier magnitude with the phase of $F^{(l)}$,

$$\hat{F}^{(l)} = |F|e^{i\angle\hat{F}^{(l)}}, \tag{5.17}$$

where $\angle\cdot$ denotes the phase of a complex number.

It remains to specify image-space constraints that define $f^{(l+1)}$ in terms of $\hat{f}^{(l+1)}$. The extent to which iterative phase retrieval is successful depends on the amount of extra information that can be incorporated in this part of the algorithm [Dainty and Fiddy 1984,Fienup and Wackerman 1986]. Examples of practical significance are : knowledge of the support of the image, positivity of the image, partially inaccurate knowledge of the scattered field phase, and partial knowledge of the magnitude of the image. Usually, at least a crude estimate of the $supp(f)$ can be inferred from $supp(ff)$. The support constraint is that $f(\vec{x}) = 0$ whenever $\vec{x} \notin supp(f)$. Positivity (see Section 2.7) is a useful theoretical constraint because it forces the image-form to be "most compact" [Bates and Fright 1985]. Denote by $\mathcal{V}^{(l)}$ the set of points at which the support constraint, and any other image-space constraints, are violated by $\hat{f}^{(l+1)}$.

The technique known as error reduction modifies $\hat{f}^{(l+1)}$ according to

$$f^{(l+1)}(\vec{x}) = 0, \quad \vec{x} \in \mathcal{V}^{(l)}, \tag{5.18}$$

$$f^{(l+1)}(\vec{x}) = \hat{f}^{(l+1)}(\vec{x}), \quad \vec{x} \notin \mathcal{V}^{(l)}. \tag{5.19}$$

An important constraint that is employed in the iterative Fienup algorithms applied to astronomical imaging is that of positivity. When positivity is invoked to determine $\mathcal{V}^{(l)}$, error reduction is known to never diverge, although stagnation frequently occurs. The hybrid input-output technique, which is defined by

$$f^{(l+1)}(\vec{x}) = f^{(l)}(\vec{x}) - \beta\hat{f}^{(l+1)}(\vec{x}), \quad \vec{x} \in \mathcal{V}^{(l)}, \tag{5.20}$$

$$f^{(l+1)}(\vec{x}) = \hat{f}^{(l+1)}(\vec{x}), \quad \vec{x} \notin \mathcal{V}^{(l)}, \tag{5.21}$$

where β is known as the relaxation parameter, usually produces better reconstructions although the error fluctuates and convergence is not assured.

In inverse scattering contexts, the equivalent source distribution is complex-valued. Consequently, the positivity constraint is meaningless. Appendix 5-A demonstrates how the performance of Fienup algorithms diminishes when the positivity constraint is removed. The only constraint imposed on the reconstructed images was a support constraint. The accuracy to which the support is known vastly affects the

performance of the algorithm [Fienup 1987]. It is interesting to note, however, that Lane [1987] shows that complex image-forms can be successfully reconstructed, but only after expenditure of so much computational effort that it is arguable whether it could ever be justified in technological contexts.

The investigation [Bates and Tan 1985b] reproduced as Appendix 5-B indicates that scattered fields recorded from time-domain inverse scattering experiments provide enough extra information to allow the scattered field phase to be readily recovered. The results demonstrate that when the extra information is incorporated into modifications of the iterative Fienup algorithms, reconstruction of complex-valued images is greatly accelerated.

A further demonstration [Gardenier *et al.* 1986] of how extra information, which is significantly erroneous, can be successfully incorporated into a phase retrieval algorithm is reproduced as Appendix 5-C. The algorithm stems from the Gerchberg-Saxton technique in electron microscopy [Gerchberg and Saxton 1972], [Saxton 1978, Chapter 5], [Gerchberg 1986].

There has recently been discovery of diffraction data that appear to exhibit forbidden crystallographic symmetries. Whilst several researchers have proposed physical models to explain this phenomenon, Appendix 5-D attempts to deduce the structure from the diffraction data alone. The algorithm converges because the image is forced to be positive.

5.4 COMPUTED TOMOGRAPHY (CT) MODELS

When the real part of χ assumes its free-space value of unity throughout all of Υ, including Υ_s, the ray paths, as given by (3.9), are merely straight lines. When two spatial dimensions are considered, it is appropriate to employ the rotated coordinate systems of Section 3.2, so that each ray is characterised by the angle ϕ between the ray direction and the fixed, unrotated y-axis, and by the perpendicular distance ξ from the ray to the origin (see Figure 5.2).

Denoting the imaginary parts of χ and the point characteristic function by f and p respectively, (3.20) reduces to

$$p(\xi;\phi) = \int f(\vec{x})\delta(\xi - \vec{x}\cdot\hat{\xi})\,d\vec{x}, \tag{5.22}$$

where $\hat{\xi}$ is the unit vector in the ξ-direction and depends on the angle ϕ. For fixed ϕ, the one-dimensional function of ξ defined by (5.22) is termed the *projection* of f at angle ϕ. The Radon transform of f is the set of its projections for $0 \leq \phi < 2\pi$ [Radon 1917,Barrett 1984]. Note that the Radon transform is a linear mapping. The ideal CT problem is to invert the Radon transform, i.e. infer f from its projections. Other names for this are computer assisted tomography, computer aided tomography, and image reconstruction from projections. McCullough and Payne [1977], Herman and Lewitt [1979], and Garden [1984] review the historical development of CT particularly informatively.

The projection theorem [Garden 1984], [Bates and McDonnell 1986, Section 9], sometimes called the projection-slice theorem [Barrett 1984], relates the one-dimensional

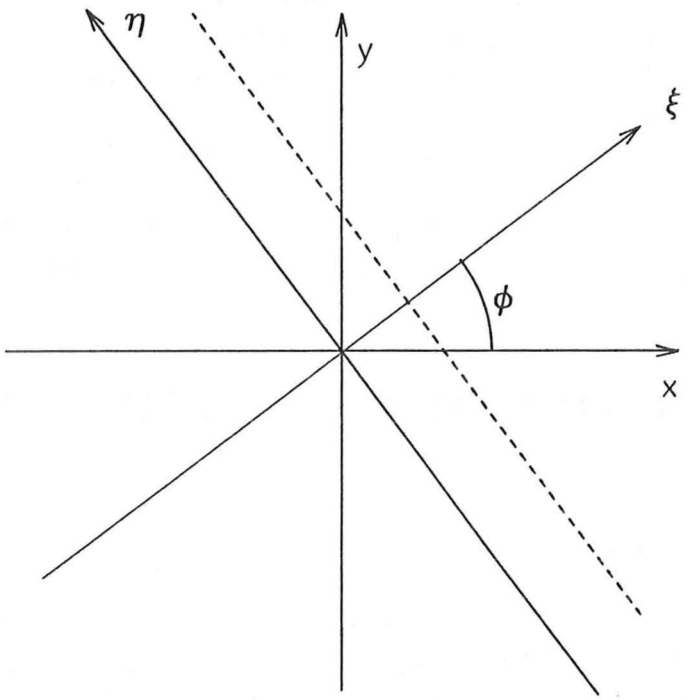

Figure 5.2: A ray is a straight line characterised by an angle ϕ and a distance ξ. The parameter along the ray is η. The value of a projection is the line integral of χ along each ray. Back-projection (see Section 6.3) of a projection assigns the value of a point on a projection to each point in the image on its associated ray.

Fourier transform of a projection to the two-dimensional Fourier transform of χ along a radial line in Fourier space :

$$\mathcal{F}_1[p(\xi;\phi)](\rho;\phi) = \mathcal{F}_2[f(r;\theta)](\rho;\phi). \tag{5.23}$$

In principle, therefore, knowledge of projections for all angles completely determines the two-dimensional Fourier transform of χ, which can itself be determined by two-dimensional inverse Fourier transformation :

$$f(r;\theta) = \mathcal{F}_2^{-1}\left[\mathcal{F}_1[p](\rho;\phi)\right](r;\theta). \tag{5.24}$$

It turns out that (5.24) does not provide the best reconstructions when the discrete nature of the data is considered. Projections, when Fourier transformed, lead naturally to samples in Fourier space that fall on a polar grid, preventing application of the FFT algorithm [Brigham 1974]. Better implementations of reconstruction from projections are discussed in Section 6.3.

Undoubtedly, the application for which the the physical process is most successfully modelled by the Radon transform is imaging of opaque objects (notably, those composed of biological tissue) with X-rays, where f is the X-ray attenuation coefficient of the body. The intensity I of an X-ray beam, initially of intensity I_0, is conventionally related to a projection by [Stanton 1969, Chapter 4], [Kak 1979], [Greenleaf 1983]

$$p(\xi;\phi) = \ln(I_0/I). \tag{5.25}$$

The advances in diagnostic radiology that have arisen from CT [Axel *et al.* 1983] are testimony to the usefulness of the Radon transform (5.22) to accurately model the propagation of the emanations employed. It should be noted, however, that the scattering of X-rays is a complicated process [Stanton 1969], [Joseph and Spital 1982] whose behaviour is not entirely faithfully described by (5.22).

The ideal projection model is so simple, because of its linearity, that one is tempted to apply it in all sorts of situations which do not strictly satisfy the assumptions on which conventional X-ray CT is predicated. The success of such a strategy depends, of course, on how large are the discrepancies between actuality and the ideal model. Some of the more pertinent deviations from the ideal are discussed below from the point of view that the data are incomplete straight-line projections.

Although the clinical and commercial success of X-ray CT in medical contexts is testimony to the reasonableness of the projection model to adequately describe the propagation of X-rays through biological tissue, it is much less appropriate in electromagnetic and acoustic applications where the physical dimensions of the inhomogeneous features are comparable to the wavelength of the radiation employed implying that the refractive index is palpably variable which invalidates the assumption that the emanations travel along straight rays. Thus, time-of-flight measurements [Johnson *et al.* 1975,Glover and Sharp 1977,Spiesberger 1985] are line-integrals along curved ray paths, as given by (3.9). Time-of-flight measurements are apparently incomplete because it is not immediately obvious how the ray paths over which the measurements are defined can be expressed in terms of any known quantities. Tomography is only effective when a sufficient number of rays traverse the cross-section being imaged. When refraction is involved, there exist "forbidden regions" [McKinnon 1980, Chapter 4], [McKinnon and Bates 1980] about which no information can be obtained

from time-of-flight measurements. There have been schemes proposed to estimate estimate straight-line projections from time-of-flight data [Norton and Linzer 1982], or to apply techniques that are modified forms of the iterative technique presented in Section 6.7 [Johnson *et al.* 1975,Schomberg 1978]. Images formed by application of similar correction schemes to integrals, along curved rays, of attenuation measurements [Kuc 1984] are more difficult to interpret because attenuation is attributable to three mechanisms : scattering, absorption and diffraction.

Even when the suitability of the projection model is accepted, there are several practical constraints that arise which cause the quantity of data to be incomplete. Practical constraints may limit the domain of each projection (with respect to ξ) [Lewitt *et al.* 1978], [Lewitt and Bates 1978b], or in angle, either evenly spaced but few [Garden 1984] or a limited angular range of projections [Sezan and Stark 1984], [Hayner and Jenkins 1984]. The definitive treatment of the limited angular range problem is by Davison [1983].

Data collected from fan-beam geometries [Bennett and Byer 1986], [Peng and Stark 1987], are essentially complete because all projections are present in the data, but in a reorganised form [Soumekh 1986b].

When considering practical numerical implementation of the inverse Radon transform, it must be remembered that only a finite number of measurements is available. It is, therefore, necessary to consider aspects of sampling and discretisation [Joseph and Schulz 1980], [O'Sullivan 1985], [Ohyama *et al.* 1987] (also see Section 1.5). The sampling theorem (refer to Section 2.7) requires samples of projections to be spaced by no more than $1/2R$ in order to reconstruct an object confined to a region of radius R. The data are incomplete if this sampling requirement is not satisfied. Although the effect of insufficient sampling of projections is qualitatively well known, it does not seem to be quantitatively appreciated the extent to which it can degrade image reconstruction. An published study to remedy this deficiency [Tan *et al.* 1986] is reproduced as Appendix 5-E.

As mentioned above, it is tempting to seek ways in which CT can be adapted to situations in which the projection model (5.22) is insufficiently representative of the physical process at hand. Other manifestations of incomplete projections include situations in which the physical process is not sufficiently well modelled. An example occurs when the generalised constitutive parameter exhibits temporal variations during scanning [Garden 1984]. A second is the phenomenon termed beam-hardening [Brooks and Di Chiro 1976a], [Alvarez and Macovski 1976], [Joseph and Spital 1978]. An X-ray beam is never perfectly monochromatic. Rather, a finite spectrum is transmitted with the lower energy (frequency) components experiencing greater attenuation. After passing through the body, the beam contains a greater proportion of higher energy X-rays and is said to be "harder". Effects of beam-hardening are most apparent at the edges of high density areas such as bone. The proposals that have been made to compensate for beam-hardening are seemingly of limited applicability because they are based on inappropriately incomplete, simplistic models of the underlying physical process [Garden 1984, Section 5.2].

5-A FOURIER PHASE RETRIEVAL WHEN THE IMAGE IS COMPLEX

Fourier phase retrieval when the image is complex

R. H. T. Bates and D. G. H. Tan

Electrical and Electronic Engineering Department,
University of Canterbury, Christchurch 1, New Zealand

Abstract

Fienup's iterative phase retrieval algorithms are applied to an image (confined to a support having a concave perimeter) whose phase can be chosen either zero, or varying arbitrarily between 0 and π, or quite arbitrary. Reconstructed images are confined (or, more precisely, we attempt to confine them) to either the support or the rectangle just enclosing it. Both positive reconstructed images are faithful (as expected) when positivity is enforced. Bipolar and complex images reconstructed within the support are recognizable, but are scarcely recognizable when reconstructed within the rectangle. The bipolar reconstructed images are superior to the complex ones. These results reinforce the general experience that a support-constraint tends by itself to be too weak to ensure faithful image reconstruction.

Introduction

Computational experience[1] now confirms that the phase of the Fourier transform of almost any two-dimensional, positive (i.e. real and non-negative), compact (i.e. of finite size, or having finite support) image can be recovered from the magnitude of its Fourier transform. Furthermore, such phase retrieval is encouragingly tolerant of noise on the given magnitude[2,3]. The samples of the magnitude must be spaced at least twice as closely as is needed to compute the image (by Fourier transformation) when the phase is also given. While much of the underlying theory has been illuminated[4,5,6] some murky aspects persist, especially as regards the effects of noise on the data[7] and the most effective way of handling images which we call foggy[8].

This paper represents a preliminary assay at rectifying the previous neglect (in the phase retrieval literature) of bipolar (i.e. real but possessing both negative and positive parts) and complex images. The rationale for hoping to be able to recover such images from their Fourier magnitudes is explained in the next section. We then outline our phase retrieval strategy. After presenting some computational results we conclude by assessing their significance and commenting on future possibilities.

Rationale for unique image recovery

Arbitrary points in image space and visibility space are identified by Cartesian coordinates (x,y) and (u,v) respectively. The image, each pixel of which can be a complex number, and its visibility are written, respectively, as $f(x,y)$ and $F(u,v)$, which satisfy

$$F(u,v) \leftrightarrow f(x,y) \tag{1}$$

where the symbol $\leftrightarrow$ identifies the members of a Fourier transform pair. Given $|F(u,v)|$ we can immediately compute (by Fourier transformation) the autocorrelation $ff(x,y)$ of the image, because it is a mathematical theorem that[6]

$$|F(u,v)|^2 \leftrightarrow ff(x,y) \tag{2}$$

We give the name image box to the region $b(x,y)$ which coincides with the support of $f(x,y)$. The corresponding region $bb(x,y)$ for $ff(x,y)$ is called the autocorrelation box. Recall that[6]

$$ff(x,y) = \iint_{b(x',y')} f^*(x',y')\, f(x'+x, y'+y)\, dx'\, dy' \tag{3}$$

where the asterisk denotes complex conjugation.

Given $|F(u,v)|$, and hence $ff(x,y)$ from (2), and hence $bb(x,y)$ from inspection of $ff(x,y)$, it is obvious from (3) that there is a lower limit on the size of $b(x,y)$. There is no upper limit because

$$|F(u,v)\, \exp(i\, \Psi(u,v))| = |F(u,v)| \tag{4}$$

where $\Psi(u,v)$ is any real function of u and v. We call any f(x,y) that exists within the smallest image box compatible with (3), for any particular ff(x,y), a most compact image. It is worth recalling that, when f(x,y) is known to be positive, the only images compatible with ff(x,y) are necessarily most compact[6].

We say that $\{\exp(i\alpha)\, f(x-a1, y-a2)\}$ and $\{\exp(i\beta)\, f^*(-x+b1, -y+b2)\}$ have the same image-form[1] as f(x,y), where α, β, a1, a2, b1 and b2 are arbitrary real constants. The image-form is all that can be recovered when phase{F(u,v)} is lost. However, each manifestation of the image-form "looks like" the image. Consequently, the image-form is all that is sought in many phase retrieval applications.

We pose the Fourier phase problem for complex images as: given |F(u,v)|, recover the most compact image-form of f(x,y).

Phase retrieval procedure

Because this paper reports only a preliminary study, we merely present results for an image having a fairly simple form (refer to next section). We therefore think it unnecessary to resort to either sophisticated initial phase estimates[1] for starting Fienup's iterative algorithms[9] or intricate sequences of the various algorithms[10]. We employ pseudo-random starting phases followed by a single error-reduction iteration, after which we execute 99 iterations of the hybrid-input-output algorithm (using a gain factor of 0.5). We thus make a direct comparison (in a special case) of the relative difficulty of reconstructing positive, bipolar and complex image-forms.

Each iteration involves a forward and a reverse Fourier transformation, which we of course implement with the FFT[11]. Sufficient samples of the visibility are included to ensure that each transformation from visibility space fills the rectangular autocorrelation box BB(x,y), where the latter's perimeter is that rectangle which just encloses bb(x,y).

The computational examples presented in the following section involve an image which can be either positive (i.e. its phase is zero), bipolar (i.e. its phase alternates arbitrarily between 0 and π) or complex (i.e. its phase is arbitrarily distributed throughout the range 0 to 2π). Positivity is enforced in image space when reconstructing the positive image-form. We apply a support-constraint in image space for all image-forms. This involves attempting to confine the reconstructed image-form to either b(x,y) or B(x,y). The latter is the rectangular image box whose perimeter is that rectangle which just encloses b(x,y). Denoting either b(x,y) or B(x,y) by Q(x,y), we recall[9] that, while the image-form is set to zero outside Q(x,y) for the error-reduction algorithm, the hybrid-input-output algorithm merely "urges" the image-form to confine itself to Q(x,y). This "urging" is effected by appropriately combining fractions of the present and previous estimates of the image-form.

The point of investigating two different "support-constraints" is that, when we have no prior knowledge of the shape of b(x,y), we may only be able to deduce B(x,y) directly from the given |F(u,v)|. It is therefore of practical interest to gain experience of how well bipolar and complex images can be reconstructed when B(x,y), rather than b(x,y), represents the support-constraint.

Computational results

We present results for an image consisting of 9 "gaussians", i.e.

$$f(x,y) = \sum_{n=1}^{9} A(n)\, \exp(-((x-\xi(n))^2 + (y-\eta(n))^2))/2\,(\sigma(n))^2) \qquad (5)$$

The values of $|A(n)|$, $\xi(n)$, $\eta(n)$ and $\sigma(n)$ are listed in Table 1, which also includes the values of phase{A(n)} for the positive, bipolar and complex versions of the image.

Table 1. Parameters for positive, bipolar and complex versions of the image

n	$\lvert A(n)\rvert$	$\xi(n)$	$\eta(n)$	$\sigma(n)$	[phase{A(n)}]/π		
					positive	bipolar	complex
1	0.33	0.00	0.00	0.10	0	0	0.33
2	0.67	0.00	0.20	0.03	0	0	1.00
3	1.00	0.30	0.20	0.03	0	1	0.25
4	0.67	0.10	0.20	0.06	0	0	1.33
5	0.33	-0.20	0.20	0.06	0	0	1.25
6	0.67	-0.10	0.15	0.03	0	1	0.00
7	0.83	-0.20	-0.15	0.03	0	1	1.80
8	0.17	-0.10	-0.20	0.06	0	1	1.00
9	0.33	0.20	-0.05	0.06	0	1	1.83

We truncate f(x,y) outside the image box b(x,y) shown in Figure 1, in which the rectangular image box B(x,y) is also shown. Contour plots of the magnitudes of the positive, bipolar and complex images are shown in Figures 2 through 4. Adjacent contours in each of these plots (and of those shown in Figures 5 through 10) correspond to magnitude changes of 0.14 of the image's maximum magnitude.

Figures 5 through 7 show the magnitudes of the images reconstructed within b(x,y). While the positive (Figure 5) and bipolar (Figure 6) reconstructions are, respectively, very and reasonably faithful, the complex reconstruction (Figure 7) is only recognizable with hindsight.

Figures 8 through 10 show the magnitudes of the images reconstructed within B(x,y). While the positive (Figure 8) reconstructions remains very faithful (as expected from general computational experience), the bipolar (Figure 9) and complex (Figure 10) reconstructions are, respectively, barely recognizable and virtually unrecognizable.

Conclusions

Our results reinforce the computational experience of everyone who has studied Fourier phase retrieval in any sort of depth: that enforcing positivity is far more effective than a support constraint. As is well appreciated now[10], it is often advantageous to relax the support constraint in the final stages of the reconstruction of a positive image.

It seems that phase retrieval suffers from a Parkinsonian principle in that reconstructed images tend to "expand to fill the space available"! They can only be dissuaded from this inflationary behavior when they are positive, because the only images then compatible with given Fourier magnitudes are necessarily most compact (as mentioned in the "rationale" section of this paper).

The accepted uniqueness arguments[4,5,6,7] are nevertheless compelling, so that we feel ourselves obliged to continue the Quest for the Grail of Direct Phase Retrieval.

Acknowledgements

One of us (D.G.H.T.) acknowledges the award of a New Zealand University Grants Committee Postgraduate Research Scholarship.

References

1. Won, M. C., Mnyama, D., Bates, R. H. T., "Improving initial phase estimates for phase retrieval algorithms," Optica Acta, in press.
2. Fienup, J. R., "Experimental evidence of the uniqueness of phase retrieval from intensity data," in Indirect Imaging (Ed. J. A. Roberts, Cambridge University Press), pp. 99-109. 1984.
3. Bates, R. H. T., Fright, W. R., Norton, W. A., "Phase restoration is successful in the optical as well as the computational laboratory," in Indirect Imaging (Ed. J. A. Roberts, Cambridge University Press), pp. 119-124. 1984.
4. Bruck, Y. M., Sodin, L. G., "On the ambiguity of the image reconstruction problem," Opt. Commun., Vol. 30, pp. 304-308. 1979.
5. Fiddy, M. A., "The phase retrieval problem," Proc. SPIE, Vol. 413, pp. 176-181. 1983.
6. Bates, R. H. T., "Uniqueness of solutions to two-dimensional Fourier phase problems for localized and positive images," Computer Vision, Graphics, and Image Processing, Vol. 25, pp. 205-217. 1984.
7. Sanz, J. L. C., Huang, T. S., "On the stability and sensitivity of multidimensional signal reconstruction from Fourier transform magnitude," Proc. I.C.A.S.S.P. (Tampa, Florida I.E.E.E. CH2118-8/85), pp. 1065-1068. 1985.
8. Bates, R. H. T., Fright, W. R., "Composite two-dimensional phase-restoration procedure," J. Opt. Soc. Am., Vol. 73, pp. 358-365. 1983.
9. Fienup, J. R., "Phase retrieval algorithms: a comparison," Appl. Opt., Vol. 21, pp. 2758-2769. 1982.
10. Bates, R. H. T., Fright, W. R., "Reconstructing images from their Fourier intensities," in Advances in Computer Vision and Image Processing (Ed. T. S. Huang, J.A.I. Press), Vol. 1, pp. 227-264. 1984.
11. Brigham, Fast Fourier Transform (Prentice-Hall. 1974).

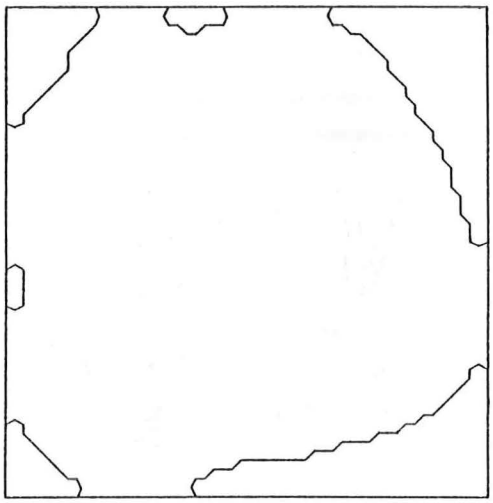

Figure 1. Image boxes. The perimeter of B(x,y) is the outer square, while the perimeter of b(x,y) is the irregular curve plus the parts of the square between points where the curve meets it.

Figure 2. Magnitude of positive f(x).

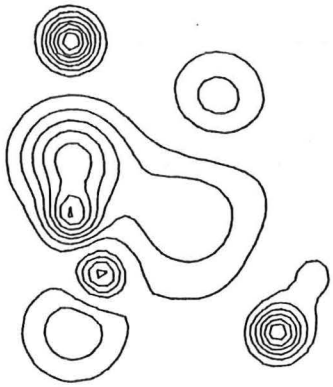

Figure 3. Magnitude of bipolar f(x).

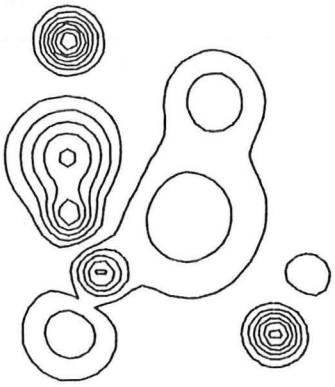

Figure 4. Magnitude of complex f(x).

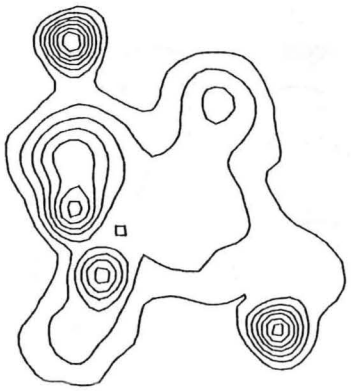

Figure 5. Magnitude of reconstruction, within b(x,y), of positive f(x).

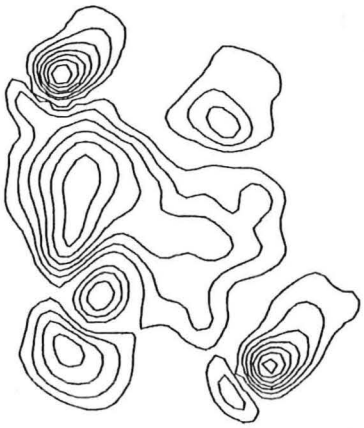

Figure 6. Magnitude of reconstruction, within b(x,y), of bipolar f(x).

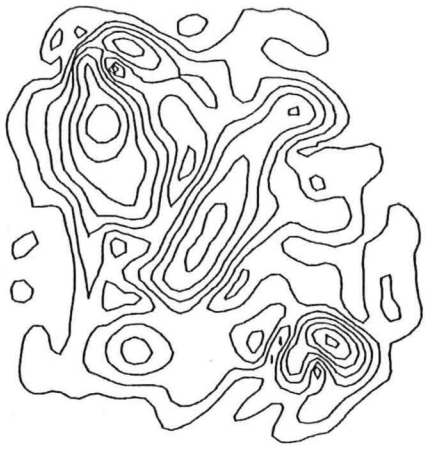

Figure 7. Magnitude of reconstruction, within b(x,y), of complex f(x).

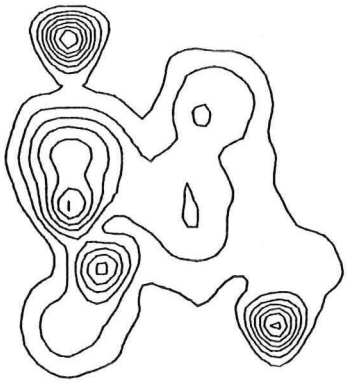

Figure 8. Magnitude of reconstruction, within B(x,y), of positive f(x).

Figure 9. Magnitude of reconstruction, within B(x,y), of bipolar f(x).

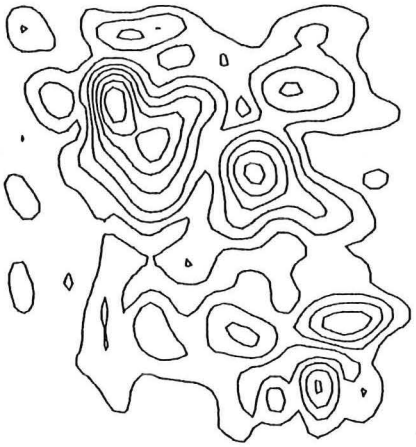

Figure 10. Magnitude of reconstruction, within B(x,y), of complex f(x).

5-B TOWARDS RECONSTRUCTING PHASES OF INVERSE SCATTERING SIGNALS

Toward reconstructing phases of inverse-scattering signals

R. H. T. Bates and D. G. H. Tan

Department of Electrical and Electronic Engineering, University of Canterbury, Christchurch, New Zealand

Received January 15, 1985; accepted July 23, 1985

Recent work on the Fourier-phase problem is shown to be relevant to the reconstruction of the relative phases of fields scattered to a large number of separated inaccurately surveyed locations. A proposed approach to phase retrieval is illustrated with a computational example. It is assumed that the waveforms of the scattered fields can be recorded at each location.

1. INTRODUCTION

Most of the algorithms[1] that have been developed for inferring the characteristics of bodies from scattered fields require the latter's phases to be known as well as their intensities. However, it is far from easy to record accurately the relative phases of electromagnetic signals received at widely separated locations. It is consequently worth inquiring whether such phases can be computed from measured intensities. We show here how recent results pertaining to the Fourier-phase problem[2] might be adapted to the recovery of inverse-scattering phases.

Pertinent features of the Fourier phase problem are summarized in Section 2. Certain technical aspects of inverse scattering are discussed in Section 3, and in Section 4 we formulate the problem of inferring the relative phases of scattered fields received at a number of separated points lying in a plane. Section 5 presents a computational example that serves both to explain our proposed approach to phase reconstruction and to demonstrate that it shows promise of being useful in practice. We assess the significance of our results in Section 6 and outline the kinds of investigation that are needed in the future.

2. CONNECTION BETWEEN FOURIER PHASE AND INTENSITY

There are two virtually distinct phase problems associated with pairs of quantities that are each other's Fourier transforms. The older of these, known as the crystallographic-phase problem,[3] cannot be attacked by the techniques summarized in this section, for a reason (both illuminating and important) that is explained later. Recent research indicates that the second problem, which we call the Fourier-phase problem, is appositely formulated as follows. First, we are presented with a positive (i.e., real and nonnegative) quantity $I(\mathbf{u})$ that is well behaved throughout an L-dimensional space (here called Fourier space) in which $\mathbf{u}$ is the position vector of an arbitrary point. It is appropriate to call $I(\mathbf{u})$ the Fourier intensity. We are then required to find that real quantity $\Phi(\mathbf{u})$, which should be associated as a phase with the positive square root of $I(\mathbf{u})$. It is appropriate to call $\Phi(\mathbf{u})$ the Fourier phase. Unless some extra condition is imposed, there are, of course, no constraints on $\Phi(\mathbf{u})$. A particularly useful way of constraining the Fourier phase is now outlined.

We define

$$|I(\mathbf{u})|^{1/2} \exp[i\Phi(\mathbf{u})] = F(\mathbf{u}) \overset{L}{\leftrightarrow} f(\mathbf{x}), \tag{1}$$

where $\overset{L}{\leftrightarrow}$ connects the members of an L-dimensional Fourier-transform pair, $\mathbf{x}$ is the position vector of an arbitrary point in a second L-dimensional space (here called image space), and $i = \sqrt{-1}$. It is appropriate to call $f(\mathbf{x})$ the image, $F(\mathbf{u})$ the Fourier complex amplitude, and $|F(\mathbf{u})| = |I(\mathbf{u})|^{1/2}$ the Fourier magnitude. We next demand that the support of the image be effectively finite, in the sense that if

$$|f(\mathbf{x})| < \epsilon \qquad \text{for all } \mathbf{x} \notin \Upsilon \tag{2}$$

where ϵ is some positive constant and Υ is a region (here called the image box) of image space, then the integral (here called the L-dimensional image volume V) over image space of the quantity, which is unity for $\mathbf{x} \in \Upsilon$ and zero for $\mathbf{x} \notin \Upsilon$, is finite. Finally, we require V to be the smallest image volume compatible with Eq. (1).

It is a mathematical theorem that[4]

$$I(\mathbf{u}) \overset{L}{\leftrightarrow} ff(\mathbf{x}), \tag{3}$$

where $ff(\mathbf{x})$ is the autocorrelation of $f(\mathbf{x})$. It is appropriate to introduce the autocorrelation box Υ_a, defined by

$$|ff(\mathbf{x})| < \eta \qquad \text{for all } \mathbf{x} \notin \Upsilon_a, \tag{4}$$

where η is another positive constant. The autocorrelation volume V_a is the integral over image space of the quantity that is unity for $\mathbf{x} \in \Upsilon_a$ and zero for $\mathbf{x} \notin \Upsilon_a$.

There is no simple connection between Υ and Υ_a, in general, when ϵ and η are chosen arbitrarily. However, when the following conditions hold (which we believe they must in the majority of physical applications of the concepts being outlined here), there is a particularly simple connection. Suppose that the form of $I(\mathbf{u})$ is such that, first, V and V_a can both remain finite, with ϵ and η being as small as we please, and, second, the surfaces σ and σ_a bounding Υ and Υ_a, respectively, are both single connected and convex. Now consider the L-dimensional rectangular parallelepipeds Λ and Λ_a that just enclose Υ and Υ_a, respectively. The sides of Λ and Λ_a are parallel to Cartesian axes set up in image space. We denote by b_l and a_l, respectively, the lengths of the sides of Λ and Λ_a in the lth direction. It is obvious from the form of the integral that defines the autocorrelation that[2]

0740-3232/85/112013-06$02.00

$$a_l = 2b_l \qquad \text{for all } l \in \{1, 2, 3, \ldots, L\}. \tag{5}$$

It is appropriate to set up Cartesian axes in Fourier space also, with the correspondence between the lth coordinate u_l and the lth coordinate x_l in image space being defined by the Fourier integral formula

$$F(\mathbf{u}) = \int_{-\infty}^{\infty} (L) \int f(\mathbf{x}) \exp(i2\pi \mathbf{u} \cdot \mathbf{x}) d\Upsilon$$

$$\text{with } \mathbf{u} \cdot \mathbf{x} = \sum_{l=1}^{L} u_l x_l, \tag{6}$$

where $\int (L) \int$ implies an L-dimensional integral and $d\Upsilon$ is the volume element in image space.

In practical applications, few physical quantities are recorded as continuous functions of time and/or distance. They are usually sampled in both time and space. The theorem represented by expression (3) thus can be taken advantage of in the real world only if the points (here called the sample points) belonging to the set $\{\mathbf{u}_j; j = 1, 2, \ldots, J\}$ at which $I(\mathbf{u})$ is given (or measured) are spaced sufficiently closely. The sampling theorem[4] demands that the spacing in the lth Cartesian direction be no greater then $1/a_l$. It is this that distinguishes the crystallographic[3] and Fourier[2] phase problems. For the former, the corresponding sample spacing is unavoidably larger, because crystallographic images are (by definition) periodic, repeated throughout image space within contiguous parallelepipeds (which are, in fact, nonrectangular, in general). When these parallelepipeds are rectangular, the Fourier intensity is observable only at points spaced by $1/b_l$ in the lth direction. Equation (5) shows this spacing to be twice $1/a_l$.

Given only the Fourier intensity, it is imposible to find where the image is located. Furthermore, one cannot distinguish between the image and the complex conjugate of its reflection in the coordinate origin. However, an object is not changed by shifting it, and its reflection remains recognizable even when its phase is reversed. We[2] say that $f(\mathbf{x} - \mathbf{y})$ and $f^*(-\mathbf{x} + \mathbf{z})$ have the same image form as $f(\mathbf{x})$, where $\mathbf{y}$ and $\mathbf{z}$ are arbitrary constant position vectors and the asterisk denotes complex conjugation.

When some Fourier phase information is available, as in the situation envisaged in Section 4, we can choose between $f(\mathbf{x})$ and $f^*(-\mathbf{x})$. The uncertainty over the image's position persists, however, but this has no significance in inverse-scattering contexts.

When $L = 1$, there is no unique connection between the Fourier intensity and the image form. When $L > 1$, however, there is almost always a unique connection (in fact, the available evidence[2] strongly suggests that, in practice, the connection is always unique for the image form whose support is as compact, as is consistent with the data).

3. INVERSE SCATTERING

Consider a finite body, say, B (refer to Fig. 1), illuminated with prescribed electromagnetic radiation. We denote a particular linearly polarized component of the scattered field (received at the point P) by the signal $s(r, \phi, \theta, t)$, where t denotes time. If, as is usual, P is in the far field (Fraunhofer region) of B for all significant frequencies present in the signal, then the r dependence becomes trivial, and the non-

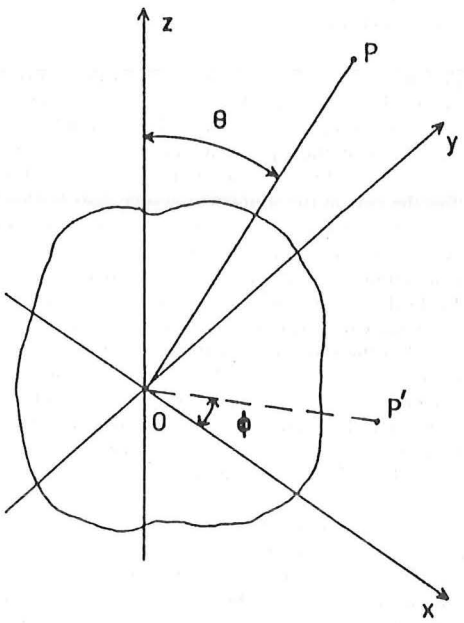

Fig. 1. Body and coordinates.

trivial functional dependence of the signal is expressed by $s(\phi, \theta, t)$, the temporal Fourier transform of which is written as $\tilde{s}(\phi, \theta, k)$, where $k = 2\pi/\lambda$ is the wave number corresponding to the free-space wavelength λ. So,

$$\tilde{s}(\phi, \theta, k) \overset{1}{\leftrightarrow} s(\phi, \theta, t). \tag{7}$$

The scattered field can be rigorously regarded as reradiations from equivalent sources, appropriately called polarization sources, initially induced in the body by the incident illumination. The reradiations induce further sources, etc., so that the actual equivalent-source distribution is self-consistent with both the illumination and the constitution of the body.[5]

When, as is usual in the real world, it is feasible to sense the scattered field only within restricted ranges (of extent appreciably less than π) of the angles ϕ and θ (we call these the angular measurement ranges), we can introduce a theoretical simplification that, although approximate, nevertheless often gives accurate results in practice. A fictitious plane (here called the equivalent-image plane) is set up contiguous to the aforesaid polarization source distribution, with the latter on the side opposite that of the sensors of the scattered field. If the correct forms are adopted for equivalent surface sources on the whole (infinite extent) plane, the far field radiated by them exactly equals $s(\phi, \theta, t)$. This is one way of expressing Huygens's principle mathematically (cf. Baker and Copson[6]).

The above-mentioned exact formulation is of little help, however, when the sensors cover only a restricted field of view. One must then resort to the following device, which is actually implicit in virtually all useful interpretations of the images formed by microscopes, theodelites, telescopes, eyes,

etc. The principle has been given precise formal expression by Silver, in his aperture-field method.[7] We draw rays from all points in the polarization source distribution to all sensors. The part of the equivalent-image plane pierced by these rays (we call this the support of the equivalent image) identifies the part of the equivalent-source distribution (we call this part the equivalent image) that is reconstructable from measurements made with the sensors. The reconstruction cannot be exact, of course, but it is quite often palpably faithful. This principle is implicit, for example, in the approximate physical-optics formulation of inverse scattering[8] and in the conceptual foundations of successful microwave-imaging techniques.[9,10] Our ultrasonic-imaging experiments[11] are explicit demonstrations of the effectiveness of this concept. One must, of course, recognize that the size of any detail that can be resolved is necessarily inversely proportional to the extents of the angular measurement ranges.

The Fourier phase problem (see Section 2) is formulated in terms of images having compact support. Since signals tend to be of palpably finite duration, it is appropriate to treat t as one image coordinate. Because the support of the equivalent image is also effectively finite, we take Cartesians (x and y, say) in the equivalent-image plane to be the other two image coordinates. To ensure that the equivalent image $f(x, y, t)$ is the Fourier transform of the scattered field, the Fourier-space coordinates must be k and[7] $u = |\sin(\theta) \times \cos(\phi)|/\lambda$ and $v = |\sin(\theta)\sin(\phi)|/\lambda$, where the pole of the spherical system (r, ϕ, θ) is to be thought of as passing through a central point of the angular measurement ranges. So x, y, and t and u, v, and k are the respective components of the three-dimensional vectors $\mathbf{x}$ and $\mathbf{u}$. We write $\tilde{s}(\phi, \theta, k)$ as $S(\mathbf{u})$.

We envisage that signals are received at points spaced in the u and v directions, respectively, by no more than one half of the reciprocal of the linear extent in the x and y directions of the equivalent image. This is necessary to permit the phase-retrieval techniques outlined in Section 2 to be applied.

Since the measurement ranges are finite, the observable $S(\mathbf{u})$ is truncated. This does not prevent imaging from being performed in practice, as explained above. It is advisable, however, to ameliorate the effects of the truncation by appropriate preprocessing, such as windowing or edge extension of the observed data[12] (the latter technique is often quite efficacious[13]). We denote the preprocessed form of $S(\mathbf{u})$ by

$$F(\mathbf{u}) = W|S(\mathbf{u})|, \tag{8}$$

where $W|\cdot|$ represents whatever form of preprocessing is adopted.

An exceedingly pertinent practical point is that $f(\mathbf{x})$ is taken to be positive (i.e., real and nonnegative) in all reports of successful phase retrieval.[2] In general, there is no reason to expect that a $F(\mathbf{u})$ pertaining to inverse-scattering data is the Fourier transform of a positive quantity. This would mean that reconstruction of inverse-scattering data would be impractical if no part of the phase of $F(\mathbf{u})$ could be measured directly. Happily, a considerable amount of phase information is immediately inferrable from measurement, as is explained in Section 4.

4. TWO-DIMENSIONAL CONSIDERATIONS

To ensure that $|F(\mathbf{u})|$ is uniquely connected with its image form, the dimension of the position vector $\mathbf{u}$ need be no more than 2. From now on we assume that P lies in the plane defined by $\phi = 0$ and $\phi = \pi$, so that the signal's functional dependence reduces to $s(\theta, t)$, and $\mathbf{u}$ has the two components u and k. The respective image-space coordinates are x and t.

Although the waveform of the signal received at P can be readily recorded for frequencies as high as those of the infrared[14] (and optical heterodyning will presumably soon become routine), there are likely to be appreciable errors in practice, however, in estimates of r and the time origin for each point P. This means that, although the major functional dependence of $\Phi(\mathbf{u}) = \Phi(u, k)$ is readily deducible from the measured data, we must expect significant errors in the average value and the slope (in the k direction) of the Fourier phase for each value of u at which data samples are available. So the estimate $\hat{\Phi}(u, k)$ of the Fourier phase obtainable directly from the measurement can be expected to be of the form

$$\hat{\Phi}(u, k) = \Phi(u, k) + \alpha(u) + k\beta(u). \tag{9}$$

The unknown $\beta(u)$ can be removed easily. We merely Fourier transform the data one-dimensionally (with respect to k) for each value of u and then shift the transformed values such that they all span the same region of t values. Consequently, the only serious errors in the data are represented by $\alpha(u)$, whose correction is described in Section 5, wherein we explain our approach to phase reconstruction with the aid of a computational example. For convenience, we take Υ to fill all Λ, whose size is normalized such that

$$b_1 = b_2 = 1 \qquad \text{so that } a_1 = a_2 = 2. \tag{10}$$

We pose the inverse-scattering phase problem as follows: Given values of $I(m/2, n/2)$ and $\hat{\Phi}(m/2, n/2)$, for sufficient integers m and n that $ff(x, t)$ can be reconstructed (to within whatever accuracy is specified) from the given samples of $I(\mathbf{u}) = I(u, k)$, reconstruct the Fourier phase $\Phi(u, k)$.

5. PHASE RECONSTRUCTION: AN EXAMPLE

The concern here is solely with the reconstruction algorithm. The data are not taken from an actual inverse-scattering problem. What needs to be demonstrated at this stage is that the Fourier phase can be readily retrieved when the image is complex. So a complex $f(\mathbf{x}) = f(x, t)$, existing for $|x| < 1/2$ and $|t| < 1/2$ in accordance with Eq. (10) has been chosen arbitrarily. The data for the inverse-scattering problem were generated from the computer Fourier transform of the image.

The first step is to deduce the size of Λ_a from the given samples of $I(u, k)$. We take this as accomplished, indicating that Λ_a has sides of length 2. As explained in Section 2, this shows immediately that the sides of Λ are of length 1.

The next step is to remove the error $k\beta(u)$ in the manner described in Section 4. Since there is no difficulty in doing this, we just assume that it is accomplished, so that the given phase can be expressed in the simplified form:

$$\hat{\Phi}(u, k) = \Phi(u, k) + \alpha(u). \tag{11}$$

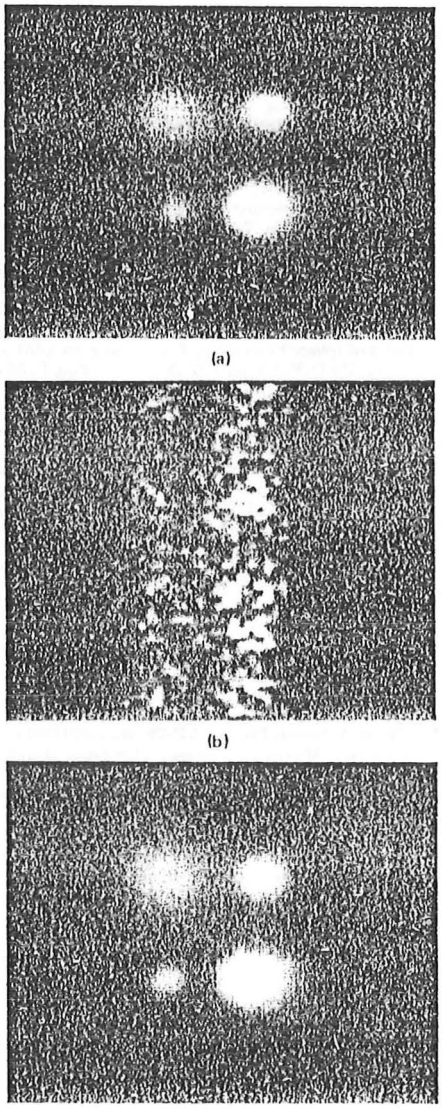

Fig. 2. Illustration of the performance of the modified Fienup algorithm: (a) $|f(\mathbf{x})|$, (b) first version of $|\tilde{f}(\mathbf{x})|$, and (c) fourth version of $|\tilde{f}(\mathbf{x})|$.

We now define $\hat{F}(u, k)$ and $\hat{f}(x, t)$ by

$$\hat{F}(u, k) = |F(u, k)| \exp(i\hat{\Phi}(u, k)) \overset{2}{\leftrightarrow} \hat{f}(x, t) = \hat{f}(\mathbf{x}), \quad (12)$$

where $|F(u, k)|$ is the positive square root of the given Fourier intensity. Since $\hat{f}(\mathbf{x})$ is computed, using the fast-Fourier-transform (FFT) algorithm,[15] from the given samples of $I(u, k)$ and $\hat{\Phi}(u, k)$, which are spaced by 1/2 in both the u and k directions, this estimate of the image has value (in general) throughout Λ_a. It is appropriate to define

$$\tilde{f}(\mathbf{x}) = \hat{f}(\mathbf{x}) \quad \text{for } \mathbf{x} \in \Lambda$$
$$= 0 \quad \text{for } \mathbf{x} \in \bar{\Lambda}_a, \quad (13)$$

where $\bar{\Lambda}_a$ is the part of Λ_a not occupied by Λ.

Writing the FFT of $\tilde{f}(\mathbf{x})$ as $\tilde{F}(\mathbf{u}) = \tilde{F}(u, k)$, we denote the phase of the latter by $\tilde{\Phi}(u, k)$.

We must now use $\tilde{\Phi}(u, k)$ to estimate the error $\alpha(u)$ in $\hat{\Phi}(u, k)$ as given by Eq. (11). For each value of u, we calculate the average difference between $\hat{\Phi}(u, k)$ and $\tilde{\Phi}(u, k)$, which we denote by

$$\Delta(u) = |1/N(u)| \sum_m |\hat{\Phi}(u, m/2) - \tilde{\Phi}(u, m/2)|, \quad (14)$$

where $N(u)$ is the number of data samples for the particular value of u (i.e., it is the number of values of k for which data are given for each value of u). We take $\Delta(u)$ to be our estimate of $\alpha(u)$. We call $[\hat{\Phi}(u, k) - \Delta(u)]$ the corrected phase.

By replacing $\hat{\Phi}(u, k)$ in Eq. (12) by the corrected phase, we set up an iterative loop, which is continued until $|\tilde{f}(x)|$ is less than some prescribed threshold throughout $\bar{\Lambda}_a$. We call this procedure the modified Fienup algorithm. The standard Fienup algorithm[2] is obtained by replacing $\hat{\Phi}(u, k)$ in Eq. (12) by $\tilde{\Phi}(u, k)$.

Figure 2(a) shows $|f(\mathbf{x})|$ for a complex image consisting of five "Gaussian blobs" whose relative phases were derived

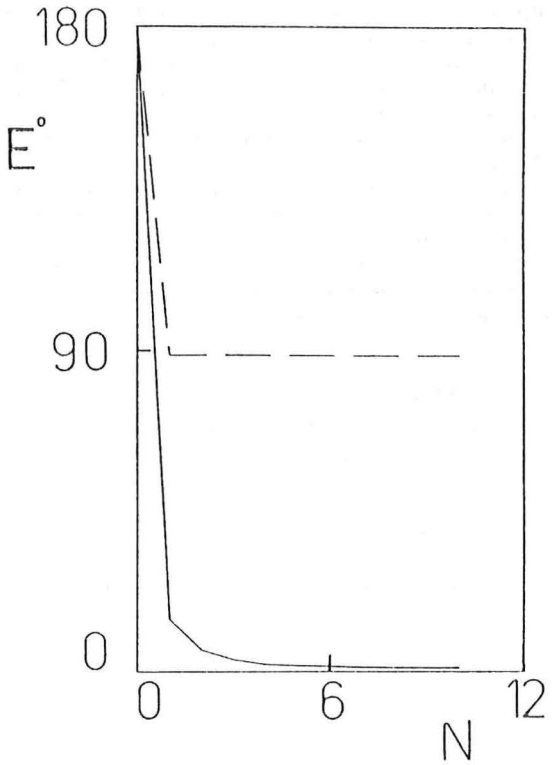

Fig. 3. Phase error E (degrees) versus number N of iterations: ——, modified Fienup algorithm; – – –, standard Fienup algorithm.

from a pseudorandom distribution uniformly distributed between 0 and 2π. Figure 2(b) shows the magnitude of the first version of $\tilde{f}(\mathbf{x})$, which bears little resemblance to the original $f(\mathbf{x})$, as is, of course, to be expected. After only four iterations, however, the form of $\tilde{f}(\mathbf{x})$ is similar to that of $f(\mathbf{x})$, as is shown in Fig. 2(c). Figure 3 shows how the mean phase error E varies with the number N of iterations, where E is the sum over all pixels in Fourier space of $|\Phi(u, k) - \tilde{\Phi}(u, k)|$ divided by the number of pixels.

The modified Fienup algorithm converges rapidly, as is seen from Fig. 3. In fact, the tenth version of $|\tilde{f}(\mathbf{x})|$ is virtually indistinguishable from $|f(\mathbf{x})|$. On the other hand, the standard Fienup algorithm does not seem to converge at all, as is confirmed in Fig. 3. For instance, at the tenth iteration the reconstructed image form is much the same as that shown in Fig. 2(b). This does not indicate any inadequacy on the part of the standard algorithm, of course, because $f(\mathbf{x})$ is complex here.

6. CONCLUSIONS

The first significant point worth noting is that the dashed curve in Fig. 3 confirms the general (although largely unrecorded, as far as the archival literature is concerned) experience of those working on Fourier-phase problems that recovery of complex image forms from their Fourier intensities is difficult. Complex images reconstructed by standard Fienup algorithms always seem to spread themselves more or less uniformly throughout A. This shows just how powerful is the constraint of "positivity," which is invoked in all previous reports of successful phase retrieval.[2]

The encouraging conclusion to be drawn from this paper is that, even when $f(\mathbf{x})$ is complex, the Fourier phase can be retrieved accurately, provided that the available estimate of it can be expressed as in Eq. (9). This suggests that it may well prove to be feasible to reconstruct the relative phases of scattered signals received at widely separated sites, thereby significantly enhancing the physical relevance of most of the existing inverse scattering algorithms.[1]

ACKNOWLEDGMENTS

We thank our colleagues W. R. Fright and D. Mnyama for their technical advice and for many helpful discussions.

REFERENCES

1. L. Colin, ed., *Mathematics of Profile Inversion*, NASA Tech. Memo TM X-62 (NASA Ames Research Center, Moffet Field, Calif., August 1972); K. Chadan and P. C. Sabatier, *Inverse Problems in Quantum Scattering Theory* (Springer-Verlag, New York, 1977); H. P. Baltes, ed., *Inverse Source Problems in Optics* (Springer-Verlag, New York, 1978); *Inverse Scattering Problems in Optics* (Springer-Verlag, New York, 1980); IEEE Trans. Antennas Propag. 29(2) (1981).
2. J. R. Fienup, "Phase retrieval algorithms: a comparison," Appl. Opt. 21, 2758–2769 (1982); R. H. T. Bates and W. R. Fright, "Composite two-dimensional phase-restoration procedure," J. Opt. Soc. Am. 73, 358–365 (1983); J. R. Fienup, in *Indirect Imaging*, J. A. Roberts, ed. (Cambridge U. Press, London, 1984); R. H. T. Bates, W. R. Fright, and W. A. Norton, in *Indirect Imaging*, J. A. Roberts, ed. (Cambridge U. Press, London, 1984); R. H. T. Bates and W. R. Fright, "Reconstructing images from their Fourier intensities" in *Advances in Computer Vision and Image Processing*, T. S. Huang, ed. (JAI, Greenwich, Conn., 1984), Vol. 1, pp. 227–264; M. C. Won, D. Mnyama, and R. H. T. Bates, "Improving initial phase estimates for phase retrieval algorithms," Opt. Acta (to be published).
3. G. N. Ramachandran and R. Srinivasan, *Fourier Methods in Crystallography* (Wiley-Interscience, New York, 1970).
4. R. N. Bracewell, *Fourier Transform and Its Applications*, 2nd ed. (McGraw-Hill, New York, 1978).
5. R. H. T. Bates and F. L. Ng, "Polarization-source formulation of electromagnetism and dielectric-loaded waveguides," Proc. Inst. Electr. Eng. 119, 1568–1574 (1972).
6. B. B. Baker and E. T. Copson, *Mathematical Theory of Huygens' Principle*, 2nd. ed. (Oxford U. Press, New York, 1953).
7. S. Silver, *Microwave Antenna Theory and Design* (Dover, New York, 1965), Chap. 5.
8. R. M. Lewis, "Physical optics inverse diffraction," IEEE Trans. Antennas Propag. AP-17, 308–314 (1969).
9. C. K. Chan and N. H. Farhat, "Frequency swept tomographic imaging of three dimensional perfectly conducting objects," IEEE Trans. Antennas Propag. AP-29, 312–319 (1981).
10. B. D. Steinberg, *Microwave Imaging with Large Antenna Arrays* (Wiley, New York, 1983).
11. R. H. T. Bates and B. S. Robinson, "Ultrasonic transmission speckle imaging," Ultrasonic Imaging 3, 378–394 (1981); R. H. T. Bates and R. A. Minard, "Compensation for multiple reflection," IEEE Trans. Sonics Ultrason. SU-31, 330–336 (1984); R. A. Minard, B. S. Robinson, and R. H. T. Bates, "Full-wave computed tomography. Part 3: coherent shift-and-add imaging," Proc. Inst. Electr. Eng. Part A 132, 50–58 (1985).
12. M. J. McDonnell, W. K. Kennedy, and R. H. T. Bates, "Identifying and overcoming practical problems of digital image restoration," N. Z. J. Sci. 19, 127–133 (1976).
13. R. M. Lewitt and R. H. T. Bates, "Image reconstruction from projections: IV: projection completion methods (computational examples)," Optik 50, 269–278 (1978); R. H. T. Bates and W. R. Fright, "Towards imaging with a speckle-interferometric optical synthesis telescope," Mon. Not. R. Astron. Soc. 198, 1017–1031 (1982).
14. E. C. Sutton, S. Subramanian, and C. H. Townes, "Interferometric measurements of stellar positions in the infrared" Astron. Astrophys. 110, 324–331 (1982); G. N. Gibson, J. Heyman, J. Jugten, W. Fitelson, and C. H. Townes, "Optical path length fluctuations in the atmosphere," preprint (Department of Physics, University of California, Calif. 94720, 1984).
15. E. O. Brigham, *Fast Fourier Transform* (Prentice-Hall, Englewood-Cliffs, N.J., 1974).

5-C APERTURE DISTRIBUTION PHASE FROM SINGLE RADIATION PATTERN MEASUREMENT VIA GERCHBERG-SAXTON ALGORITHM

APERTURE DISTRIBUTION PHASE FROM SINGLE RADIATION PATTERN MEASUREMENT VIA GERCHBERG–SAXTON ALGORITHM

Indexing terms: Antennas, Antenna aperture distributions, Antenna radiation patterns, Phase retrieval

Computer simulations indicate that antenna aperture phase distributions can be obtained, by invoking the Gerchberg–Saxton algorithm appropriately modified, from a single far-field radiation pattern, only the magnitude of which is measured. The pattern must be sampled at significantly more than the Nyquist rate, and a sequence of initial phase estimates must be tried.

The radiation patterns of large steerable antennas in general, and ground-based antennas for satellite communications systems in particular, can be measured with the aid of special satellites.[1] It is hardly practical to determine radiation pattern phase directly unless some holographic scheme is employed.[2,3] The latter can be inconvenient because it requires a separate large 'reference' antenna to be placed near the antenna under test.

If phase $\{F(u)\}$ is known as well as $|F(u, v)|$, where $F(u, v)$ is an antenna's far-field radiation pattern, with u and v being the usual spatial frequency co-ordinates,[4] then the aperture distribution $f(x, y)$ can be immediately computed by Fourier transformation, where x and y are the aperture plane Cartesian co-ordinates conjugate to u and v, respectively. Inspection of phase $\{f(x, y)\}$ reveals the locations of imperfections in a reflector surface and/or displacement of a feed. Morris and others[5,6] have shown it to be practical to recover phase $\{f(x, y)\}$ from radiation pattern magnitudes measured with the feed in two positions along the antenna axis. Recently Anderson and Sali[7] have reported a new algorithm which uses two separate measurements of the magnitude of the field measured in the Fresnel region. These authors imply that pairs of radiation patterns need to be measured if one is to recover the aperture phase reliably.

Since it is often impractical to make accurate measurements at two different ranges less than the far-field distance, and because it is inconvenient to displace the feed of a large antenna (unless necessary to upgrade antenna performance), we have been studying whether the Gerchberg–Saxton algorithm[8,9] can be appropriately modified so as to provide a really practical means of inferring the required information from a single far-field radiation pattern measurement.

We envisage that samples of $|F(u, v)|$ are obtained for spatial frequencies $u = m\alpha/A$ and $v = n\alpha/B$, for as many values of the integers m and n as span the spatial frequency range for which $|F(u, v)|$ is significant, where A and B are the sides of the rectangle which just encloses the aperture. Our crucial finding, inspired by work on the Fourier phase problem,[10,11] is that the factor α should be appreciably less than the value unity specified by the sampling theorem.[12] There is no technical difficulty involved in ensuring that $\alpha < 1$.

The standard Gerchberg–Saxton algorithm requires both $|F(u, v)|$ and $|f(x, y)|$ to be given. Although the latter is unknown to us in practice, a reasonable approximation to it is its ideal (i.e. designed) form denoted here by $g(x, y)$.

We begin by choosing (in the manner described below) an initial aperture phase here denoted by $\psi(x, y)$, which is attached to the ideal $g(x, y)$, giving

$$\hat{f}(x, y) = g(x, y) \exp(j\psi(x, y)) \tag{1}$$

as the first estimate of $f(x, y)$. The Fourier transform (always computed with the aid of the FFT algorithm) of $\hat{f}(x, y)$ is $\hat{F}(u, v)$. The phase of the latter is attached to the measured $|F(u, v)|$ to give the first estimate of the complex radiation pattern. This is then inverse-Fourier-transformed and the phase of the result replaces $\psi(x, y)$ in eqn. 1, thereby setting up an iterative loop of Gerchberg–Saxton type. The degree to which the algorithm has converged, at each iteration, is assessed by the error E defined by

$$E = \int_R (|\hat{F}(u, v)| - |F(u, v)|)^2 \, du \, dv \tag{2}$$

where R is the region of the u, v-plane spanned by the measured samples of $|F(u, v)|$.

Fig. 1 illustrates the convergence of this algorithm when $\psi(x, y)$ is generated by a pseudorandom number routine and $|f(x, y)|$ is identical to $g(x, y)$ (i.e. the standard Gerchberg–Saxton situation). The solid and broken curves in Fig. 1 apply for $\alpha = 1/2$ and $\alpha = 1$, respectively. The increase in the rate of convergence for the smaller value of α is marked. While accelerated numerical convergence is obviously desirable, it is of minor practical significance of itself, because the cost of the needed computer time is negligible compared with the overall monetary value of a satellite communication system (or even of a radio telescope). The real reason for reducing the value of α becomes fully apparent when one simulates the kinds of aperture imperfections that are likely to be of practical importance. We then find that it is often impossible to achieve convergence unless α is made appreciably less than unity.

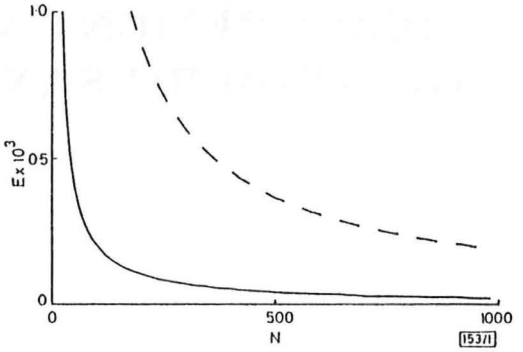

Fig. 1 *E against number N of iterations for $f(x, y) = g(x, y) = 1$ over a circular aperture*

Both curves are averages for 10 separate simulations, each starting with a different pseudorandom $\psi(x, y)$:
——— $\alpha = 1/2$
- - - - $\alpha = 1$

Our experience suggests that convergence is always achieved, with $\alpha = 1/2$ and $\psi(x, y)$ made pseudorandom, when $|f(x, y)|$ differs pseudorandomly from $g(x, y)$ and phase $\{f(x, y)\}$ is pseudorandom. A greater challenge to the algorithm is presented when the aperture phase contains a 'defocus' term corresponding to the feed being displaced. It is then necessary for $\psi(x, y)$ to correspond to a 'guessed' degree of defocus (here represented by the parameter θ, which is the phase error at the edge of the aperture, due to defocus).

Fig. 2 shows the variation of E with the number N of iterations for various values of θ, with $\alpha = 1/2$, when $\phi = 10$ was the actual degree of defocus. In this simulation, phase $\{f(x, y)\}$ also contained a term, of Gaussian form, which we inserted so as to model a 'dent' in the reflector. The dent was centred halfway between the middle and the edge of the aperture, and its effective diameter was 1/8 that of the aperture. To provide a severe test for the algorithm we chose the maximum phase error due to the dent to be 5 rad. We took $g(x, y) = 1$ within the aperture. Pseudorandom noise was added to the radiation pattern so that, at each far-field pixel, $|F(u, v)|$ differed from its ideal value by up to $|F(0, 0)|/20$. For no value of θ was convergence achieved when we took $\alpha = 1$.

Inspection of Fig. 2 indicates that convergence is only achieved when θ is sufficiently close to ϕ. The practical implication of this is that one must keep on trying different values of θ until numerical convergence becomes manifest. The procedure would be computationally intensive, but would nevertheless be economically viable, for the reasons stated earlier and because there is no need to make any physical adjustment

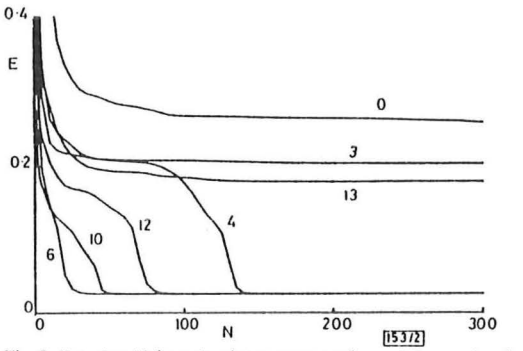

Fig. 2 *E against N for a circular aperture, with $\alpha = 1/2$, $g(x, y) = 1$*

Aperture phase contains terms corresponding to dent in reflector (see text) and displaced feed ($\phi = 10$), and $F(u, v)$ taken as Fourier transform of aperture distribution corrupted with additive pseudorandom noise (see text). Number assigned to each curve is value of θ

to the antenna while gathering the data on which the algorithm operates.

We are currently studying how to increase the efficiency of our modified Gerchberg–Saxton algorithm by incorporating Fienup's[13] and our own[10,11] improved phase retrieval procedures.

P. H. GARDENIER
C. A. LIM
D. G. H. TAN
R. H. T. BATES

Electrical & Electronic Engineering Department
University of Canterbury
Christchurch 1, New Zealand

28th November 1985

References

1 GREEN, H. E.: 'Antenna pattern measurement with a geostationary satellite', *J. Electr. & Electron. Eng., Aust.*, 1983, 3, pp. 8–17
2 ANDERSON, A. P.: 'Microwave holography', *Proc. IEE*, 1977, **124**, No. 11R, *IEE Rev.*, pp. 946–961
3 BENNETT, J. C., ANDERSON, A. P., MCINNES, P. A., and WHITAKER, A. J. T.: 'Microwave holographic metrology of large reflector antennas', *IEEE Trans.*, 1976, **AP-24**, pp. 295–303
4 SILVER, S.: 'Microwave antenna theory and design' (Dover, New York, 1965), Section 6·3
5 MORRIS, D.: 'Phase retrieval in the radio holography of reflector antennas and radio telescopes', *IEEE Trans.*, 1985, AP-33, pp. 749–755
6 ELLDER, J., LUNDAHL, L., and MORRIS, D.: 'Test of phase-retrieval holography on the Onsala 20 m radiotelescope', *Electron. Lett.*, 1984, 20, pp. 709–710
7 ANDERSON, A. P., and SALI, S.: 'New possibilities for phaseless microwave diagnostics. Part 1: Error reduction techniques', *IEE Proc. H, Microwaves, Antennas & Propag.*, 1985, **132**, pp. 191–198
8 GERCHBERG, R. W., and SAXTON, W. O.: 'A practical algorithm for the determination of phase from image and diffraction plane pictures', *Optik*, 1972, 35, pp. 237–246
9 SAXTON, W. O.: 'Computer techniques for image processing in electron microscopy' (Academic Press, New York, 1978), Chap. 5·3
10 BATES, R. H. T., and FRIGHT, W. R.: 'Composite two-dimensional phase-restoration procedure', *J. Opt. Soc. Am.*, 1983, **73**, pp. 358–365
11 WON, M. C., MNYAMA, D., and BATES, R. H. T.: 'Improving initial phase estimates for phase retrieval algorithms', *Opt. Acta*, 1985, 32, pp. 377–396
12 BRACEWELL, R. N.: 'The Fourier transform and its applications' (McGraw-Hill, New York, 1983, 2nd edn.), Chap. 10
13 FIENUP, J. R.: 'Phase retrieval algorithms: a comparison', *Appl. Opt.*, 1982, **21**, pp. 2758–2769

5-D PHASE RETRIEVAL FOR ELECTRON DIFFRACTION PATTERNS SUGGESTING ORIENTATIONAL ORDER WITHOUT TRANSLATIONAL SYMMETRY

Optik 73, No. 2 (1986) 69–73 © Wissenschaftliche Verlagsgesellschaft mbH, Stuttgart

Phase retrieval for electron diffraction patterns suggesting orientational order without translational symmetry

R. H. T. Bates and D. G. H. Tan

Electrical and Electronic Engineering Department, University of Canterbury, Christchurch, New Zealand

Received 31 October 1985

Abstract

Results are presented of applying Fourier phase retrieval techniques to recently reported electron diffraction patterns from specimens of rapidly cooled Al/Mn alloy. The structure is assumed positive, since phase retrieval would otherwise be impossible. Diffraction patterns suggestive of five-fold rotational symmetry lead to structures which exhibit that symmetry approximately and which appear to be dominated by discrete "blobs". When the symmetry is enforced, the blob structure becomes more pronounced, in accord with recent theoretical predictions. What may be unexpected is that the blobs have a wide range of amplitudes.

Inhalt

Wiederherstellung der Objektphase aus Elektronenbeugungsbildern, die auf Rotationssymmetrie ohne Translationssymmetrie hindeuten. Es werden Ergebnisse zur Wiederherstellung der Objektphase aus Elektronenbeugungsbildern mit der Fouriertechnik am Beispiel der schnell abgekühlten Al/Mn-Legierung dargestellt. Damit eine Phasenrekonstruktion möglich ist, wird die Materialstruktur als positiv angenommen. Beugungsbilder, die auf eine 5zählige Rotationssymmetrie hindeuten, führen zu Strukturen, die diese Symmetrie ungefähr aufweisen und welche durch diskrete „blobs" beherrscht erscheinen. Bei erhöhter Symmetrie werden diese „blobs" ausgeprägter, was auch in neuen theoretischen Vorhersagen angedeutet wird. Unerwartet sind die großen Amplitudenbereiche der „blobs".

1. Introduction

We demonstrate, by example, that recently developed Fourier phase retrieval techniques [1, 2, 3] can be successfully applied to the fascinating electron diffraction patterns (consisting of discrete spots suggestive of crystal structures) that have lately been observed from certain rapidly cooled alloys of aluminum [4]. These patterns are inconsistent with classical crystals, however, because their rotational symmetries do not accord with structures possessing translational symmetries. Such patterns, which have simultaneously and seemingly coincidentally appealed to theoreticians [5, 6], are beginning to excite wide scientific interest [7].

Because of the potentially deep implications of the aforesaid patterns, it seemed to us that it would be useful to attempt phase retrieval making as few initial assumptions as possible. The point is that one would like to know whether the observed patterns are consistent with physically reasonable structures significantly different from those envisaged by the theoreticians [5, 6]. The particular patterns suggestive of five-fold rotational symmetry were what especially intrigued us since we thought (for the reasons given in § 2) that they might contain sufficient information to allow our image reconstruction methods [2, 3] to be immediately applied.

In § 2 we establish necessary notation, explain why it should be feasible to retrieve the phase of particular patterns which interest us, and list the assumptions that cannot be avoided if phase retrieval is to be attempted. The diffraction pattern information available to us [4, 5] had to be pre-processed, in the manner described in § 3, before we could embark on phase retrieval. The image reconstructed on the basis of the minimum practicable number of assumptions is presented in § 4, while in § 5 we indicate how this image changes when further *a priori* assumptions are made. In § 6, we assess the significance of our results.

2. Notation and rationale

We take Cartesian coordinates u, v in the plane in which an electron diffraction pattern $F(u, v)$ is observed. We call the two-dimensional Fourier transform of $F(u, v)$ the image $f(x, y)$, where x and y are image-plane Cartesian coordinates conjugate to u and v, i.e.

$$F(u, v) = \iint_{B(x,y)} f(x, y) \exp\left[i2\pi(ux + vy)\right] dx\, dy \quad (1)$$

where $B(x, y)$, which we call the image box, is the region of the image plane within which the image exists, i.e.

$$f(x, y) = 0 \text{ for } (x, y) \text{ outside } B(x, y) \quad (2)$$

If only the intensity $|F(u, v)|^2$ of the pattern is observed, an estimate of its phase must somehow be generated before the image can itself be reconstructed by invoking the inverse of the transformation defined by (1).

There is no *a priori* necessity to place a constraint on the phase of $f(x, y)$. The latter can be complex (i.e. its phase can vary arbitrarily from point to point in image space), unlike a crystallographic image which must be positive (i.e. real and non-negative) [10].

Fourier phase retrieval [1, 2, 3] can at present only be performed successfully when the image is positive. We cannot yet handle complex images [8], unless much additional information is available [9]. The diffraction pattern intensities in which we are interested (e.g. the ones presented by Shechtman et al. [4] that are suggestive of five-fold rotationally symmetric images) are themselves ten-fold rotationally symmetric, so that they are certainly consistent with real images. Mainly because we would otherwise be unable to proceed, we make the stronger assumption that the images *are* positive. In mitigation we plead that this additional assumption has theoretical support [5, 6].

Our phase retrieval methods [2. 3] require $|F(u, v)|$ to be oversampled compared with what is necessary when the phase of $F(u, v)$ can also be observed. The X-ray crystallographers make no such restriction, but their techniques [10] rest on the knowledge that the image is constituted of atoms, whose individual scattering characteristics are well understood. Furthermore, many of the crystal structures which are presently considered most significant cannot be unravelled from a single X-ray diffraction pattern (it is often essential to resort to isomorphous replacement, etc.). We do not need any *a priori* information concerning the component parts of the image, nor do we ever demand more than a single diffraction pattern.

When we first saw the above-mentioned ten-fold symmetric diffraction patterns, it occurred to us that the density of the discrete diffraction spots might be high enough to represent a significant degree of oversampling. The results reported in Sections 4 and 5 indicate that this conjecture is at least reasonable.

Table 1. Details of diffraction spots taken from 1a of Levine and Steinhardt [5]. The 25 spots listed here are confined to a single 36° sector of u, v-space. So, each listed spot, except for the spot centred at the origin, is repeated in each of the other nine sectors, giving 241 spots in all. The numbering is purely for identification purposes. The relative brightness of each spot is proportional to the diameter of the corresponding circle on Levine and Steinhardt's figure. The u, v coordinates are normalised.

Spot Number	u coordinate	v coordinate	Relative brightness
1	0.00	0.00	1.00
2	0.00	0.29	0.24
3	0.00	0.48	0.59
4	0.00	0.77	0.82
5	0.00	0.96	0.12
6	0.00	1.26	0.94
7	0.00	1.56	0.47
8	0.00	1.75	0.35
9	0.17	0.54	0.18
10	0.17	1.01	0.29
11	0.17	1.50	0.24
12	0.17	1.81	0.24
13	0.28	0.86	0.53
14	0.28	1.17	0.41
15	0.28	1.35	0.12
16	0.28	1.65	0.65
17	0.45	0.92	0.29
18	0.45	1.11	0.41
19	0.45	1.41	0.76
20	0.45	1.90	0.82
21	0.56	1.26	0.12
22	0.74	1.31	0.18
23	0.74	1.49	0.65
24	0.74	1.79	0.71
25	0.92	1.56	0.18

3. Computational preliminaries

We do not know of any quantitative listings of the measured diffraction patterns. Informed opinion [7, 11] seems to be, however, that Levine and Steinhardt's calculated patterns [5] are effectively identical to those which have been observed [4, 11]. Levine and Steinhardt's fig. 1a depicts the diffraction spots as circles, whose diameters are directly proportional to the relative brightnesses of the spots. So, we examined Levine and Steinhardt's figure with a graduated eyepiece, and table 1 of this paper shows the result of our labours. We refer to "brightness" in table 1 because we were at first unclear as to whether Levine and Steinhardt's term "intensity" implies $|F(u, v)|$ or $|F(u, v)|^2$. We eventually decided on the latter, after making the check described in the final paragraph of this section.

The diffraction spots are ideally two-dimensional delta functions, and there is in principle an infinite number of them. Our phase retrieval techniques require that $f(x, y)$ and $F(u, v)$ each be effectivley zero outside a finite region (which we call a box) in, respectively, the x, y-plane and the u, v-plane. So we had to pre-process the ideal diffraction pattern to force it to be consistent with these constraints.

Denoting the ideal diffraction pattern by $G(u, v)$, we generate from it another pattern, written as $E(u, v)$, which is the two-dimensional Fourier transform of an image, written as $e(x, y)$, confined to an effectively finite box in the image plane. We do this by convolving the delta function representing each diffraction spot with a Gaussian function $\exp[-2(u^2 + v^2)\,\pi^2 r^2]$, where r is the effective radius of the above image box. Note that

$$e(x, y) = (1/2\pi r^2)\, g(x, y) \exp[-(x^2 + y^2)/2r^2] \qquad (3)$$

where $g(x, y)$ is the ideal image, which is the two-dimensional Fourier transform of $G(u, v)$.

The next step is to generate the $F(u, v)$, to which we actually applied our phase retrieval algorithms, by windowing $E(u, v)$ with another Gaussian function:

$$F(u, v) = E(u, v) \exp[-(u^2 + v^2)/2R^2] \qquad (4)$$

where R is the effective radius of the Fourier box (i.e. the region of the u, v-plane to which $F(u, v)$ is effectively confined). We chose the value R such that only the 241 spots implied by table 1 are manifest in $|F(u, v)|$, which is shown in fig. 1.

The convolution theorem [12] reminds us that the windowing of $E(u, v)$ defined by (4) implies that $f(x, y)$ is the convolution of $e(x, y)$ with the two-dimensional Fourier transform of $\exp[-(u^2 + v^2)/2R^2]$. This implies that the effective radius of the image box is increased by the factor $[1 + (1/4\pi^2r^2R^2)]^{1/2}$, which is in fact little different from unity because $(2\pi rR)$ is the ratio of the effective radii of the Fourier box and each of the "Gaussian spots" depicted in fig. 1.

By Fourier transforming $|F(u, v)|^2$ we obtain the image's autocorrelation [12], which must be positive if $f(x, y)$ is positive. Similarly, the Fourier transform of $|F(u, v)|^4$ is the autocorrelation's autocorrelation, which must of course also be positive. This suggested the following way of checking whether the brightnesses listed in table 1 are proportional to the amplitude $|G(u, v)|$ or the intensity $|G(u, v)|^2$ of the ideal pattern. Denoting the amplitude of the pre-processed diffraction pattern shown in fig. 1 by $A(u, v)$, the peaks of each of whose spots are proportional to the square root of the corresponding brightness listed in table 1, we Fourier transformed $A(u, v)$, $A^2(u, v)$ and $A^4(u, v)$. Whereas the first of these transforms possessed negative parts of appreciable magnitude, the latter two transforms were wholly positive. This persuaded us that $A(u, v)$, and hence fig. 1, and the square roots of the brightnesses listed in table 1 are proportional to $|F(u, v)|$.

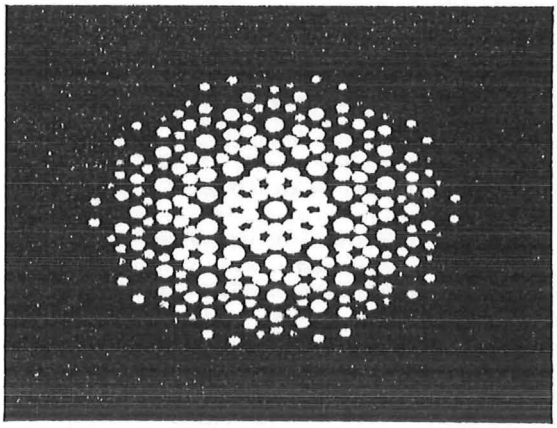

Fig. 1. $|F(u, v)|$ generated, as described in § 3, from the square roots of the brightnesses of the diffraction spots listed in table 1.

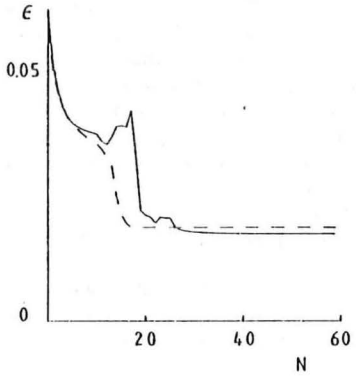

Fig. 2. Error curves (ε versus N):
- - - - - - only positivity enforced
——— both positivity and rotational symmetry enforced

4. Basic reconstructed image

We sampled fig. 1 to form a representation of $|F(u, v)|$ consisting of 128 × 128 pixels. We then assigned pseudo-random phases to each diffraction spot, insisting that phases of spots reflected in the origin of the u, v-plane must be equal but opposite, thereby ensuring that $f(x, y)$ is real. The overall phase retrieval strategy [1, 2, 3] gradually forces the image to become positive. The phase is retrieved iteratively by Fourier transforming back and forth between the image plane and the u, v-plane. Before transforming to the former from the latter, we each time force the same phase on every pixel belonging to any particular spot. We do this because the spots are discrete. Consequently, at each iteration, there are as many individual phases as there are spots (with the phases grouped in equal but opposite pairs, as indicated in the second sentence of this paragraph).

Positivity is the only constraint we can enforce in the image plane. Accordingly we define the image-error ε, after each transformation from the u, v-plane, as the sum of the magnitudes of all negative pixels divided by $F(0, 0)$. The latter necessarily equals the sum of all pixels in the image once it has become entirely positive. This ideal goal can never be reached, of course, because of imperfections of the data and round-off computational errors, and because the data may indeed be incompatible with complete positivity. Nevertheless, the variation of ε with the number N of iterations is the only available objective criterion of the effectiveness of our phase retrieval procedure.

Two of Fienup's [1] iterative algorithms are especially useful: the error-reduction algorithm, which expunges all the negative parts of the image at each iteration, and the hybrid-input-output algorithm, which merely "urges" positivity on the image (described in detail elsewhere [2]). The former algorithm is best invoked to begin with. It tends to stagnate relatively, quickly, however, after which the hybrid-input-output algorithm comes into its own.

The dashed curve in fig. 2 shows the variation of ε with N for 10 error-reduction iterations followed by 10 hybrid-input-output iterations. The initial rapid decrease of ε, after the hybrid-input-output is introduced at $N = 11$, together with the slow increase when N reaches 18, is

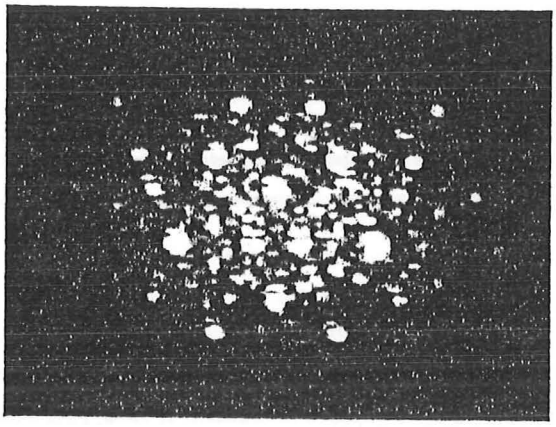

Fig. 3. Best image obtained after enforcing only positivity.

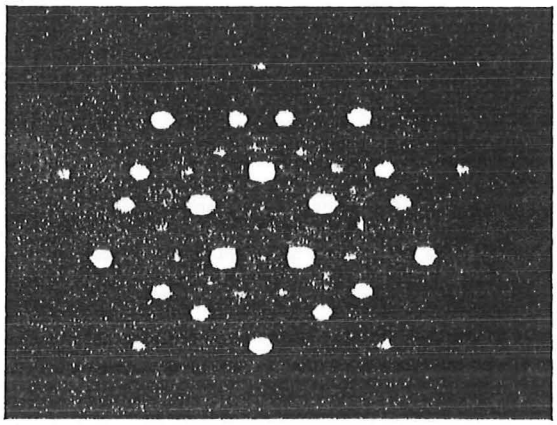

Fig. 4. Best image obtained after enforcing both positivity and five-fold rotational symmetry.

typical of this algorithm (it is worth noting that the error-reduction algorithm must converge, which is of course why it also tends to stagnate). The minimum of the dashed curve is at $N = 18$. Fig. 3 shows the form that the image has acquired by then. It is dominated by discrete blobs, in accord with the theory [5, 6].

Careful inspection of fig. 3 suggests that the phase retrieval procedure is trying to force the image to be five-fold rotationally symmetric. The centre of rotation seems to be displaced towards the upper middle of the image. Further hybrid-input-output iterations were seen to degrade the clarity of the blobs, but also moved the centre of rotation down towards the centre of the array. We follow up this deduction in the next section.

5. Consequences of assuming five-fold rotational symmetry and discrete point form for the image

When $f(x, y)$ is five-fold rotationally symmetric then so must $F(u, v)$ be. The diffraction spots in fig. 1 are arranged in groups of ten, each member of any group being at the same radial distance from the origin of the u, v-plane and separated in angle from its adjacent fellows by $\pi/5$. Having assigned a pseudo-random phase, ψ say, to one member of a group, its adjacent fellows are assigned the phase $(-\psi)$, and their adjacent fellows are assigned $(+\psi)$, etc. This scheme, which is enforced at each iteration, maintains realness and five-fold rotational symmetry of the image.

The solid curve in fig. 2 shows ε versus N for 10 error-reduction iterations followed by 15 hybrid-input-output iterations and another 35 error-reduction iterations. We are unsurprised (because of our extensive experience of Fourier phase retrieval) by the solid curve being above the lowest level reached by the dashed curve until $N = 28$, because the enforcement of rotational symmetry slows numerical convergence until the Fourier phase approaches quite close to its "true" value, after which the symmetry helps to make ε decrease more rapidly.

Table 2. Details of primary blobs apparent in fig. 4. Each blob is repeated at angular spacings of $\pi/5$ radians. The adjusted amplitude is the amplitude of the spot in fig. 4 divided by the Gaussian factor $(1/2\pi r^2)\exp[-(x^2+y^2)/2r^2]$, which is introduced in (3).

Primary blob number	Radius	θ	Adjusted Amplitude
1	2.10	0.32	1.87
2	3.29	0.74	0.19
3	3.37	1.17	0.21
4	4.69	0.14	1.16
5	4.74	0.48	1.17
6	5.70	0.95	3.20
7	6.99	0.32	1.68

Stagnation sets in at around $N = 48$ (i.e. the solid curve in fig. 2 hardly decreases between $N = 48$ and $N = 60$). The form the image takes after the 48th iteration is shown in fig. 4, which is seen to be even more dominated by discrete blobs than fig. 3. The blobs of course form groups of 5. Table 2 lists the "adjusted amplitudes" and cylindrical polar coordinates (radius, θ) of the more pronounced blobs, numbered according to increasing radial coordinate. Only the "primary blob" of each group is listed (this is the blob whose angular coordinate, within the range, $0 \leq \theta < 2\pi$, is least). Each amplitude is adjusted by dividing the amplitude of the corresponding blob in fig. 4 by $(1/2\pi r^2)\exp[-(x^2+y^2)/2r^2]$, which is the Gaussian factor introduced in (3) to ensure that the image is confined to an effectively finite region of image space. Since the theoreticians [5, 6] seem to envisage a distribution of equal-amplitude blobs, table 2 permits their ideas to be directly compared with our results.

6. Discussion

Our most significant result is perhaps that we have demonstrated that phases can be objectively and systematically retrieved for certain electron diffraction patterns from particular specimens of aluminum alloys measured by Shechtman et al. [4]. We do not have to make any structural assumptions. When only positivity is enforced, successive images appear to be tending towards being five-fold rotationally symmetric, so that we feel justified in enforcing this symmetry explicitly. We have no means of assessing whether the faint background in fig. 4 (i.e. the part of the image not comprised of pronounced blobs) is structurally significant or has the character of numerical "round-off error". We cannot prove that our reconstructed image is "unique", but virtually all previous computational experience of Fourier phase retrieval (summarised in [1, 2, 3]) suggests that it is (at least within the limits set by any round-off errors).

The low values to which the error curves in fig. 2 fall indicate, first, that our assumption of the image being positive is certainly reasonable and, second, that the reconstructed image must be close to being as fully refined as the data allow.

Being neither crystallographers nor material scientists we are not competent to comment on the physical significance of fig. 4. We leave that to those (cf. [13, 14]) who are currently arguing the meaning of Shechtman et al's. [4] delightful experiment.

We thank our colleague (in the Department of Chemistry) Ward T. Robinson for making us aware of this interesting development in crystallographic/electron-diffraction science.

References

[1] J.R.Fienup, Appl. Optics 21 (1982) 2758.
[2] R.H.T.Bates, W.R.Fright, in Advances im Computer Vision and Image Processing 1 (T.S.Huang ed.). JAI Press, Greenwich, Connecticut (1984) chapter 5.
[3] M.C.Won, D.Mnyama, R.H.T.Bates, Optica Acta 32 (1985) 377.
[4] D.Shechtman, I.Blech, D.Gratias, J.W.Cahn, Phys. Rev. Lett. 53 (1984) 1951.
[5] D.Levine, P.J.Steinhardt, Phys. Rev. Lett. 53 (1984) 2477.
[6] P.Kramer, R.Neri, Acta Crystallographica A 40 (1984) 580.
[7] BMS: Physics Today 38 (Feb. 1985) 17.
[8] R.H.T.Bates, D.G.H.Tan, Proc. SPIE 558 (1985).
[9] R.H.T.Bates, D.G.H.Tan, J. Opt. Soc. Amer. 2 A (1985) 2013.
[10] G.Ramachandran, R.Srinivasan, Fourier Methods in Crystallography. Wiley N.Y. (1970).
[11] P.A.Bancel, P.A.Heiney, P.W.Stephens, A.I.Goldman, P.M.Horn, Phys. Rev. Lett. 54 (1985) 2422.
[12] R.N.Bracewell, Fourier Transform and its Applications. McGraw-Hill, N.Y. (1978) 2nd edn.
[13] A.L.Mackay, P.Kramer, Nature 316 (1985) 17.
[14] L.Pauling, Nature 317 (1985) 512.

5-E IMAGE RECONSTRUCTION FROM PROJECTIONS: VIII: EFFECTS OF FINITE RESOLUTION AND SAMPLING OF INDIVIDUAL PROJECTIONS

Optik 73, No. 1 (1986) 25–29 © Wissenschaftliche Verlagsgesellschaft mbH, Stuttgart

Image reconstruction from projections: VIII: Effects of finite resolution and sampling of individual projections

D. G. H. Tan, J. X. Qu*) and R. H. T. Bates

Electrical and Electronic Engineering Department, University of Canterbury, Christchurch, New Zealand

Received 22 October 1985

Abstract

Effects of finite resolution and spacing of the detectors, with which projections are measured, are evaluated with reference to a model image consisting of a pair of "Gaussian blobs". The major effects are aliasing and loss of resolution and contrast. Formulas which describe the effects quantitatively are illustrated with selected computed curves.

Inhalt

Bildrekonstruktion aus der Projektion: VIII. Einfluß der endlichen Auflösung und Abtastung der einzelnen Projektionen. Der Einfluß der endlichen Auflösung und des Abstandes der Detektoren mit denen die Projektionen gemessen werden, wurden bezüglich eines Modellbildes, das aus einem Paar „Gaußscher Blobs" besteht, berechnet. Die bedeutenden Einflüsse sind Geisterbilder und Verluste in Auflösung und Kontrast. Formeln für die quantitative Beschreibung der Effekte werden mit ausgewählten Computerkurven illustriert.

1. Introduction

This is the eighth in a series of papers [1, 2, 3, 4] dealing with image reconstruction from imperfect projections. Our concern here is with projections sampled with finite resolution at discrete points. So far as we have been able to determine, the literature relating to these important questions is surprisingly sparse [5, 6, 7, 8, 9], which is why we have devoted effort at this comparatively late date to studies that one might have expected to have been disposed of some time ago.

In § 2 we list necessary preliminaries. The effect of any real-world measurement being inevitably of finite resolution is analysed in § 3. The spatial sampling of projection measurements is examined in § 4, while the further filtering needed to reduce aliasing is formulated in § 5. In order to derive widely applicable results, whose significance can be readily appreciated, we need a simple image model, which is introduced in § 6. Quantitative consequences of finite measurement resolution are studied in § 7, and in § 8 we present a number of graphs which characterise the effects on reconstructed images of finite sampling and of filtering in Fourier space. In § 9 we assess the significance of our results.

2. Preliminaries

Three-dimensional images of a body are conveniently built up from two-dimensional images of parallel cross-sections through the body. The circle C (of radius normalised to unity), shown in fig. 1, circumscribes a particular cross-section which we wish to image. The particular quantity, which we are required to make images of, is here called the density, denoted by $\lambda(r, \theta)$, where r and θ are cylindrical polar coordinates corresponding to the Cartesian coordinates x and y shown in fig. 1:

$$x = r\cos(\theta) \quad \text{and} \quad y = r\sin(\theta) \tag{1}$$

The x, y-coordinates are considered to be fixed in the cross-section whereas the ξ, η-coordinates can rotate with respect to them by the arbitrary angle φ. The data, on which we are to operate with the goal of forming an image of the density, are constituted from what we call rays. The particular ray, at distance ξ and angle φ, is defined by

$$f(\xi, \varphi) = \int_{-\infty}^{\infty} \lambda(r, \theta)\, d\eta \tag{2}$$

where the infinite limits merely emphasise that the ray passes through the whole cross-section. Inspection of fig. 1 reveals that these limits can always be replaced by $\pm (1 - \xi^2)^{1/2}$ if one so desires (it is, however, usually more convenient to write the limits as $\pm\infty$).

When the data consist of all rays, within the interval $|\xi| < 1$, we call $f(\xi, \varphi)$ the projection at angle φ. The projection theorem states that [1]

$$f(\xi, \varphi) \leftrightarrow (1)\, \Lambda(\varrho, \varphi) \tag{3}$$

where

$$\lambda(r, \theta) \leftrightarrow (2)\, \Lambda(\varrho, \varphi) \tag{4}$$

*) on leave from:
Department of Radio and Electronics, University of Science and Technology of China, Hefei, Anhui, the People's Republic of China.

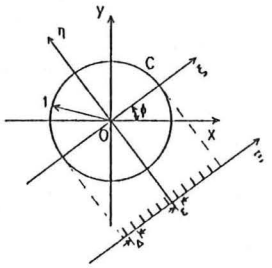

Fig. 1. Coordinates and notation appertaining to a cross-section through a body.

where $\leftrightarrow(n)$ identifies quantities forming an n-dimensional Fourier transform pair. The cylindrical polar coordinates ϱ and φ identify an arbitrary point in Fourier space. For $f(\xi, \varphi)$ and $\Lambda(\varrho, \varphi)$ to be singlevalued, we see that

$$f(-\xi,\varphi)=f(\xi,\varphi+\pi) \text{ and } \Lambda(-\varrho,\varphi)=\Lambda(\varrho,\varphi+\pi) \quad (5)$$

The conventional way to reconstruct images is by modified backprojection [2, 10]. The latter is, however, theoretically equivalent to combining (3) and (4), which are more convenient to invoke when one is making comparative studies of the kind with which we are here concerned. We take such a "Fourier approach" to image reconstruction throughout this paper.

3. Finite resolution

It is impossible to measure a single ray in practice. A practical measurement, denoted here by $p(\xi, \varphi)$, is more realistically represented by the one-dimensional convolution of $f(\xi, \varphi)$ with the instrument function, or point spread function (psf), of whatever detects the physical process manifesting the rays. We write the psf as $h(\xi)$, so that

$$p(\xi, \varphi) = f(\xi, \varphi) \odot (1)\, h(\xi) \quad (6)$$

where $\odot(n)$ is the n-dimensional convolution operator. On defining

$$p(\xi, \varphi) \leftrightarrow (1)\, \Omega(\varrho, \varphi) \quad (7)$$

and appealing to (3), we see from the convolution theorem [11] that

$$\Omega(\varrho, \varphi) = H(\varrho)\, \Lambda(\varrho, \varphi) \quad (8)$$

where $h(\xi) \leftrightarrow (1)\, H(\varrho)$.

The presence of $H(\varrho)$ on the right hand side (RHS) of (8) emphasises that the resolution of the measurement is set by the range of spatial frequencies inherent in the psf. It therefore makes sense to call $p(\xi, \varphi)$ a blurred ray, or a blurred projection when it is given throughout the range $|\xi| < 1$.

4. Sampling

In few practical applications are measurements made continuously in ξ. The blurred projections are more often sampled. We now examine the effect of sampling at the discrete points $\xi = (\varepsilon + m\Delta)$, where m is any integer and ε is called the offset. These sampling points are indicated in fig. 1 on the Ξ-axis, which is parallel to the ξ-axis. It is informative to think of a linear array of detectors with their apertures tangent to the Ξ-axis, the centre of each aperture coincident with one of the sampling points. We call Δ the sampling interval.

We write the sampled blurred projection at angle φ as

$$sp(\xi, \varphi) = p(\xi, \varphi)\, s(\xi) \leftrightarrow (1)\, S\Omega(\varrho, \varphi) \quad (9)$$

where the sampling function $s(\xi)$ is defined by

$$s(\xi) = \sum_{m=-\infty}^{\infty} \delta(\xi - \varepsilon - m\Delta) \leftrightarrow (1)\, S(\varrho) \quad (10)$$

where $\delta(\cdot)$ denotes the Dirac delta function. The infinite limits on the summation in (10) are inserted solely for notational convenience, like the infinite limits on the integral in (2). The range of m need of course only be such that all sampling points within $|\xi| < 1$ are included. Since [11]

$$S(\varrho) = \exp(i2\pi\varrho\varepsilon) \sum_{m=-\infty}^{\infty} \delta(\varrho\Delta - m) \quad (11)$$

we can again appeal to the convolution theorem to deduce from (7) and (9) that

$$S\Omega(\varrho, \varphi) = \Omega(\varrho, \varphi) \odot (1)\, S(\varrho)$$
$$= (1/\Delta) \sum_{m=-\infty}^{\infty} \exp(i2m\pi\varepsilon/\Delta)\, \Omega(\varrho - m/\Delta, \varphi) \quad (12)$$

We see that $S\Omega(\varrho, \varphi)$ consists of $\Omega(\varrho, \varphi)$ infinitely repeated throughout Fourier space, centred on circles of radius zero, $1/\Delta$, $2/\Delta$, etc., as depicted in fig. 2.

5. Fourier filtering

Unless we take appropriate precautions, the image reconstructed from $S\Omega(\varrho, \varphi)$ will exhibit an unacceptable level of aliasing. We must take steps to inhibit all terms in the summation in (12), except that for $m = 0$.

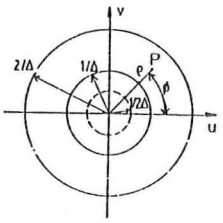

Fig. 2. Coordinates and notation appertaining to Fourier space.

Inspection of fig. 2 suggests, perhaps, that we might do best to reject all of $S\Omega(\varrho, \varphi)$ for $\varrho > 1/2A$. This is satisfactory if A is small enough that $\Omega(\varrho, \varphi)$ is negligible for $\varrho > 1/2A$. When A is larger than this, however, simple truncation of $S\Omega(\varrho, \varphi)$ is liable to lead to unacceptable ripples (or "ringing" artefacts) running across the reconstructed image. It is then appropriate to introduce a filter, $G(\varrho, \varphi)$ say, which is unity at $\varrho = 0$ and remains nearly so until ϱ approaches $1/2A$, after which $G(\varrho, \varphi)$ falls smoothly to zero. Such a filter represents a compromise between aliasing and further resolution loss. The shape of the filter affects the level of ringing artefacts in the reconstructed image (the critical consideration is that the shape should be smooth).

We write the Fourier-filtered version of $S\Omega(\varrho, \varphi)$ as

$$FS\Omega(\varrho, \varphi) = G(\varrho, \varphi)\, S\Omega(\varrho, \varphi) \tag{13}$$

The convolution theorem then confirms that the reconstructable image, denoted by $r\lambda(r, 0)$, is given by

$$r\lambda(r, 0) = s\lambda(r, 0) \odot (2)\, g(r, 0) \tag{14}$$

where $r\lambda(r, 0) \leftrightarrow (2)\, FS\Omega(\varrho, \varphi)$, $s\lambda(r, 0) \leftrightarrow (2)\, S\Omega(\varrho, \varphi)$ and $g(r, 0) \leftrightarrow (2)\, G(\varrho, \varphi)$.

It is important to recognise that only a limited number of (blurred) projections are measured in practice. The actually reconstructable image is thus even more imperfect than $r\lambda(r, 0)$, as defined by (14). However, since its effects can be separately assessed, because they are now well understood [1, 10], we feel it is unnecessary to take explicit account here of the number of projections being finite.

6. Model image

Since image reconstruction based on (3) and (4), or equivalently modified back-projection [2], is linear in the data and the image, we can assess the effects of finite resolution and of sampling by appropriate analysis of simple models. The particular model image which we choose to examine here is

$$\lambda(r, 0) = \exp(-r^2/2b^2) + A \exp[-(r^2 + \tau^2 - 2\tau r \cos(0 - \psi))/2a^2] \tag{15}$$

where the positive, real constants a, A, b, τ and ψ are chosen to be such that $\lambda(r, 0)$ is negligible for $r > 1$. Note that we have a "two-gaussian model", with a and b being the "effective radii" of the two "blobs" of density constituting the model. We always take

$$b > a \tag{16}$$

so that the wider blob is centred at the origin. We call A the amplitude of the narrower blob, which is centred at the point whose polar coordinates are (τ, ψ).

Even though many different types of detector are employed in practical applications [5], we assert that a Gaussian psf can characterise the main effects of finite resolution. The point is that it is the "effective width", σ say, of $h(\xi)$ that chiefly determines how "blurred" $p(\xi, \varphi)$ is compared with $f(\xi, \varphi)$. So we take

$$h(\xi) = \exp(-\xi^2/2\sigma^2) \tag{17}$$

It is usually appropriate to make the Fourier filter $G(\varrho, \varphi)$ circularly symmetric. Since a gaussian shape satisfies most practical "smoothness" constraints, we feel it is satisfactory to set

$$G(\varrho, \varphi) = G(\varrho) = \exp(-2\pi^2 D^2 \varrho^2) \tag{18}$$

It is the ratio of D to A which governs the compromise (noted in § 5) between aliasing and further resolution loss.

7. Contrast loss due to finite resolution

It is first of all informative to examine the reconstructable model image without concerning ourselves with the effects of sampling. It then transpires that $r\lambda(r, 0)$ is the two-dimensional Fourier transform of $\Omega(\varrho, \varphi)$, as defined by (8). On invoking (15) and (17), and evaluating the various Fourier integrals, and normalising the amplitude of $r\lambda(r, 0)$ by multiplying it by $(1/(2\pi)^{1/2}\, \sigma)$, we see that

$$r\lambda(r, 0) = B' \exp(-r^2/2\beta^2) + A' \exp[-(r^2 + \tau^2 - 2\tau r \cos(0 - \psi))/2\alpha^2] \tag{19}$$

where

$$B' = b^2/(b^2 + \sigma^2),\ \beta = (b^2 + \sigma^2)^{1/2},$$
$$A' = Aa^2/(a^2 + \sigma^2),\ \alpha = (a^2 + \sigma^2)^{1/2} \tag{20}$$

It is useful to take the first term on RHS (15) to represent a "background" spreading across the whole cross-section, thereby implying that b is of the order of unity. Interesting images cannot be formed if the resolution of the measurements is only comparable to b. We must require that

$$\sigma \ll b \tag{21}$$

in which case B' and β effectively reduce to unity and b respectively. So, the wider blob is effectively unchanged in the reconstructed image. Comparing the second terms on RHS (15) and RHS (19), we see that the amplitude and effective radius of the narrower blob are reconstructed as A' and α respectively.

Inspection of (20) indicates that whether or not the reconstructed image of the narrower blob is degraded depends upon the relative values of a and σ. When the narrower blob is well below the resolution limit of the measurement (i.e. $a \ll \sigma$), it transpires that α and A'/A reduce to σ and $(a/\sigma)^2$ respectively. We call

$$A'/A = a^2/(a^2 + \sigma^2) \tag{22}$$

the contrast loss due to finite resolution.

8. Effects of sampling and filtering

It is clear that useful images can only be reconstructed if

$$\Delta \lesssim b \tag{23}$$

When this condition is combined with (21), we find that reconstructions of the wider blob are negligibly degraded by sampling and Fourier filtering, provided $G(\varrho, \varphi)$ is itself negligible for $\varrho > (1/\Delta + w)$, where w is such that the two-dimensional Fourier transform of $\exp(-r^2/2b^2)$ is negligible for $\varrho > w$. Since this Fourier transform is necessarily "narrow", because the wider blob is by definition "wide", there is no practical difficulty in maintaining the constraint on $G(\varrho, \varphi)$. It then transpires that the part of $r\lambda(r, 0)$, as defined by (14), relating to the wider blob is given to good approximation by $((2\pi)^{1/2}\,\sigma/\Delta)\exp(-r^2/2b^2)$. Consequently, provided $r\lambda(r, 0)$ is normalised by multiplying it by $(\Delta/(2\pi)^{1/2}\,\sigma)$, we need pay no more attention to the wider blob.

The above discussion of the wider blob implies that the blurring and contrast loss of the reconstructed narrower blob (noted in § 7) are independent of the latter blob's position in the cross-section. So, for convenience, we henceforth set

$$\tau = 0 \tag{24}$$

On adopting the above normalisation, and maintaining the same constraints on b, we find from (11), (12), (13), (24) and the definitions introduced in § 6 that

$$\begin{aligned}(1/2\pi)^2\, FS\Omega(\varrho, \varphi) &= \exp(-2\pi^2 b^2 \varrho^2) \\ &+ A\, a^2 \exp(-2\pi^2(a^2 + \sigma^2 + D^2)\,\varrho^2) \\ &+ A\, a^2 \sum_{m=1}^{\infty} \{\exp[-2\pi^2(a^2+\sigma^2)(\varrho - m/\Delta)^2 + i2m\pi\varepsilon/\Delta] \\ &+ \exp[-2\pi^2(a^2 + \sigma^2)(\varrho + m/\Delta)^2 - i2m\pi\varepsilon/\Delta]\} \\ &\qquad \exp(-2\pi^2 D^2 \varrho^2)\end{aligned} \tag{25}$$

It is the summation in (25) which characterises the effects of finite sampling. Whether or not sampling introduces significant distortion can be gauged from the relative amplitudes, at $\varrho = 1/2\Delta$, of the second exponential term and the first term in the summation. The fraction of the integral over Fourier space, for $\varrho > 1/2\Delta$, of the second exponential term is, in fact, directly proportional to its value at $\varrho = 1/2\Delta$. So, the factor af (here called the aliased fraction), defined by

$$af = \exp(-\pi^2(a^2 + \sigma^2 + D^2)/2\Delta^2), \tag{26}$$

represents an estimate of the integrated intensity of the fraction of the reconstructed image which is aliased (i.e. that fraction which represents the distortion of the reconstructed image caused by finite sampling).

The two-dimensional Fourier transform of $FS\Omega(\varrho, \varphi)$, as defined by (25), can be expressed in terms of complex error functions, but it is as convenient computationally to perform the integration numerically. We denote by $a\lambda(r)$ the reconstruction of the narrower blob obtained by two-dimensionally Fourier transforming the terms in (25) which are multiplied by A. Fig. 3 shows the variation of $a\lambda(r)$ with r for $A = 1$, $\varepsilon = 0$, $a/\sigma = 0.2$ and $\Delta/\sigma = 1.5$, for several values of Δ/D. When D is small, $a\lambda(r)$ is seen to be significantly distorted from its true gaussian shape. As D is increased, the distortion becomes less pronounced, but there is a loss of both resolution and contrast, in that $a\lambda(r)$ is both broader and of lower amplitude than $\exp(-r^2/2a^2)$.

The effective radius of the reconstructed narrower blob is always close to $(a^2 + \sigma^2 + D^2)^{1/2}$, which is the effective radius of the twodimensional Fourier transform of the second exponential term in (25), throughout the ranges of the practical interest of the parameters a, σ, D, ε and Δ. Therefore, the factor

$$rl = (1 + (\sigma/a)^2 + (D/a)^2)^{1/2} \tag{27}$$

is a meaningful measure of the resolution loss.

The contrast loss, here denoted by cl, is given by A'/A, where A' is the amplitude at $r = 0$ of the Fourier transform of the terms in (25) which are multiplied by A. We find that

$$\begin{aligned}cl = (a^2/(k^2 + D^2))\,\{1 + 2 \sum_{m=1}^{\infty} [&\exp(-(k^2 + D^2)\, q_m^2) \\ &+ \pi^{1/2} k q_m \exp(-D^2 q_m^2)\,\mathrm{erf}(k q_m)]\cos(2m\varepsilon\pi/\Delta)\}\end{aligned} \tag{28}$$

where $k^2 = (a^2 + \sigma^2)$, $q_m^2 = (2\pi^2 m^2 k^2/(k^2 + D^2)\,\Delta^2)$ and $\mathrm{erf}(x) = (2/\pi^{1/2}) \int_0^x \exp(-t^2)\, dt$.

The formula (28) for cl is readily evaluated. We give a few specimen results here for illustrative purposes. Fig. 4 shows cl versus Δ/D for $\varepsilon = 0$ and $a/\Delta = 1.0$, for several values of σ/Δ. To demonstrate the combined effects of finite sampling and a finite offset, we show in figs. 5 and 6 the variation of cl with Δ/D and σ/Δ, respectively, for $\varepsilon/\Delta = 0.4$ and for various combinations of parameters.

Points particularly worthy of notice are that cl reduces to the value of A'/A given by (22) when Δ becomes small,

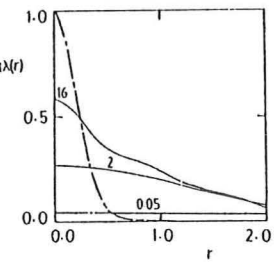

Fig. 3. Profiles of reconstructions of narrower blob for $A = 1$, $\varepsilon = 0$, $a/\sigma = 0.2$, and $\Delta/\sigma = 1.5$. Solid curves are profiles for indicated values of Δ/D. Dashed curve is ideal profile $\exp(-r^2/2a^2)$.

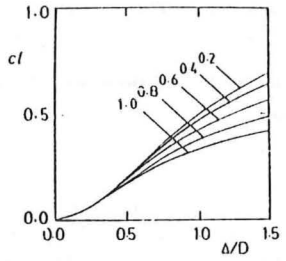

Fig. 4. Contrast loss versus Δ/D for $\varepsilon = 0$ and $a/\Delta = 1.0$. Curves show cl for indicated values of a/Δ.

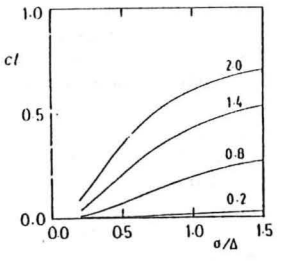

Fig. 5. Contrast loss versus Δ/D for $\varepsilon/\Delta = 0.4$ and $\sigma/D = 1.6$. Curves show cl for indicated values of a/σ.

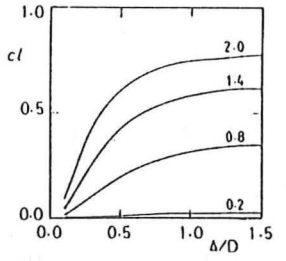

Fig. 6. Contrast loss versus a/Δ for $\varepsilon/\Delta = 0.4$ and $\Delta/D = 0.8$. Curves show cl for indicated values of a/σ.

and cl decreases markedly when both a/Δ and σ/Δ are appreciably less than unity and $(a/\Delta) < \varepsilon/\Delta < (1 - a/\Delta)$. In the latter case, the detectors are comparatively insensitive to the narrower blob.

9. Conclusions

The formulas presented in this paper permit the conquences of finite resolution and sampling to be quantitatively assessed simply and validly. We assert that these results should permit particular reconstructed images, such those presented by other authors [6, 9], to be more easily assessed objectively.

The formulas (26) and (27), for aliasing fraction af and resolution loss rl, respectively, are simple enough to be calculated by hand. The formula (28) for contrast loss cl must be computed digitally, but it could be evaluated conveniently enough, in any particular instance, for ranges of interest of the several parameters.

The simple formula (22) for contrast loss, valid when Δ is small, is quite significant, for a reason pointed out to us by G. Kossoff of the Ultrasonics Institute in Sydney, Australia. When employing computed tomography (of any kind, e.g. ultrasonic, X-ray, low-frequency electric current, etc. [10]) for routine screening purposes, it could be both effective and economically advantageous to employ a comparatively small number of relatively wide, but contiguous, detectors. Image features narrower than the detectors could still be imaged successfully, with negligible aliasing (because of the closeness of the detectors) but of course at poorer resolution. A vital consideration is, however, that the features must stand out adequately from their backgrounds, because they may otherwise be unrecognisable in reconstructed images, owing to the contrast loss which is given by (22).

One of us (D.G.H.T) acknowledges the award of a Postgraduate Research Scholarship from the New Zealand University Grants Committee, while another (J.X.Q) thanks the University of Science and Technology of China, for providing financial support and the Ministry of Foreign Affairs of New Zealand, for making necessary arrangements for his visit.

References

[1] R.M. Lewitt, R.H.T. Bates, Optik **50** (1978) 19, 189, 269.
[2] R.M. Lewitt, R.H.T. Bates, T.M. Peters, Optik **50** (1978) 85.
[3] P.B. Heffernan, R.H.T. Bates, Optik **56** (1980) 101 u. **60** (1982) 85.
[4] K.L. Garden, R.H.T. Bates, Optik **68** (1984) 161.
[5] B.K.P. Horn, Proc. IEEE **66** (1978) 551.
[6] C.R. Crawford, A.C. Kak, Appl. Optics **18** (1979) 3704.
[7] H. Stark, J. Opt. Soc. Am. **69** (1979) 1519.
[8] P.A. Rattey, A.G. Lindgren, IEEE Trans. **ASSP-29** (1981) 994.
[9] R.A. Brooks, G.H. Glover, A.J. Talbert, R.L. Eisner, F.A. Dibianca: J. Compt. Assist. Tomogr. **3** (1979) 511.
[10] R.H.T. Bates, K.L. Garden, T.M. Peters, Proc. IEEE **71** (1983) 356.
[11] R.N. Bracewell, The Fourier Transform and Its Applications, New York, McGraw-Hill (1978) 2nd edn.

Chapter 6

INVERSION ALGORITHMS

6.1 INTRODUCTION

This chapter discusses algorithms for solving inverse problems that are complete in the sense that the quantity and quality of the data available is sufficient to permit inversion to be accomplished. Even when an inverse problem is complete, there remain major conceptual and computational difficulties. It is shown in Chapter 4 that Ω is in general a nonlinear mapping of C, which almost always makes it impossible to devise an analytical representation for the inverse operator. One has to resort to iterative approaches which, although they may be ingenious, tend to be *ad hoc*. Furthermore, it is only rarely that one can answer any of the associated uniqueness questions. Such solutions to inverse problems are by no means useless, however, because they can be assessed (and often further refined) in the light of previous practical experience.

When the connection between E and C is so simple that Ω is linear, it is usually easy to construct the inverse transformation Ω^{-1} [Lee 1984]. Uniqueness questions can be readily answered and the sensitivity of the reconstructed generalised constitutive parameter to both errors in the data, and to numerical inaccuracies in the implementation, is straightforwardly assessable [Oldenburg 1984]. This is highlighted by the many successful applications of the principles of echo-location, computed tomography (CT), and Fourier imaging which are routinely invoked nowadays throughout science and engineering. Even in situations where these principles are not strictly applicable, they are often invoked, with results that (while inferior to those obtained when the implied assumptions are valid, or at least nearly so) are often far from completely useless (refer to Table 1.1).

The reasoning presented in Section 6.2 places the inverse scattering problem in the context of the coherent inverse source problem, and highlights the practical computational difficulties in the way of obtaining numerical solutions to inverse problems. The point here is that the dimensionality of the data is necessarily less than that of the quantities which one wishes to recover from the data. It is worth contrasting this with direct problems which can always (at least in principle) be cast as Fredholm integral equations implying that the dimensionality of the unknown (the scattered field in this context) is no greater than that of the supplied information. For inverse problems, the dimensionality of the data and the unknowns can only be made equal by introducing fairly sweeping approximations. As indicated in Section 6.2, this does not mean that solutions cannot be obtained unless such approximations are invoked, but it does emphasise why so little real progress has been made in solving many of the inverse

problems which have been studied intensively over the last quarter century.

The remaining sections of this chapter represent a hierarchy of approaches to generating numerical solutions to complete inverse problems. The approximations that are invoked are less sweeping, and the analytical and computational complexities increase accordingly, with each section until Section 6.7, where no explicit approximations are invoked.

In Section 6.3, wave motion is approximated by emanations described in terms of straight rays which exhibit negligible spreading as they propagate away from the source. Such approximations are highly reasonable (in the sense that they lead to very useful results) in a number of applications. The archetypal instance is X-ray computed tomography (CT) in which the quantity to be reconstructed is the X-ray attenuation coefficient of a body that has been probed [Kak 1979,Bates *et al.* 1983]. Although the conventional projection model introduced in Section 5.4 leads notionally to a procedure for inferring the generalised constitutive parameter from samples, supplied on a radial grid, of its Fourier transform, the generally preferred method (which is described in Section 6.3) is modified, or filtered, back-projection.

Because X-ray CT has revolutionised medical diagnostic radiology (see, for example, Herman [1979]), it is appropriate to examine any theoretical, experimental, and computational aspects of CT that deviate from the ideal CT model. Some of these have already been discussed in Chapter 5 (see Section 5.4). Note also that Appendix 5-E is relevant in this regard.

Section 6.4 considers another deviation from the ideal CT situation. The sources of radiation are still assumed to give rise to emanations that propagate as straight rays but, rather than being controlled by the experimenter, are embedded within the body. It is the spatial distribution of the emitting substance that is to be inferred. Employment of gamma cameras to detect emission of photons from radionuclides deposited within the human body was commonplace before X-ray CT came on the scene, but the images so obtained were merely equivalent to conventional X-ray shadowgram projections [Garrett and Smithson 1987,Keyes 1987]. True cross-sectional reconstruction was only attempted after the advent of CT. Whilst the interaction of the emanations and the body are well understood on the atomic scale, it is far from clear how to accurately model the macroscopic process. The emphasis here is on finding a model that is sufficiently physically representative whilst still permitting the construction of tractable computational inversion algorithms. The nature of the interaction between the emanations and the body has two important facets that should be taken into account. The first is that the emanations suffer attenuation while passing through the body. Secondly, the detectors employed have appreciable beamwidths. Appendix 6-A is apparently the first investigation to employ a model simultaneously incorporating both facets. As in almost all other previous studies [Tanaka *et al.* 1984], a constant attenuation coefficient is assumed (for the reasons given in Appendix 6-A).

When diffraction effects are significant, a description in terms of rays becomes inadequate. The approximations outlined in Sections 4.2.1 and 4.2.2 provide linear expressions incorporating diffraction. As shown in Section 6.5, the far-field form of these approximations lead to tractable inversion schemes, known as diffraction tomography [Mueller 1980,Devaney 1982] or Fourier imaging [Robinson 1982, Chapter 1].

Section 6.6 is devoted to the compensation for what is now generally known as "speckle" [Dainty 1975a,Abbott and Thurstone 1979]. It is manifested whenever

quasi-monochromatic wave motion, such as laser light [Erf 1978] or narrow-band ultrasound [Burckhardt 1978,Bates and Robinson 1981], is scattered from or transmitted through an inhomogeneous propagation medium. It is also generated by phase instabilities in coherent receiving systems [Robinson 1982]. The particular kind of speckle processing known as shift-and-add is examined from the point of view of reducing the effects on the reconstructed image of multiple reflections that are incorrectly interpreted by the simple model employed in echo-location. There is no lack of data in the situations discussed in Section 6.6. The recorded data are, however, "jumbled up". Shift-and-add can be regarded as a type of preprocessing that "unjumbles" the data.

The use of approximate inverse scattering models has been successful in practice, partially due to the relative simplicity of the inversion schemes, which are invariably manifested as explicit non-iterative algorithms. There are always, however, situations in which the errors inherent in assuming an approximate model become intolerable, requiring a more faithful nonlinear model to be adopted. Explicit inversion of a general nonlinear relation is impractical. However, iterative inversion can be contrived, as indicated in Section 6.7, which outlines the general Newton-Kantorovich technique for solving nonlinear equations. The method has been previously applied in various forms but it seems not to have been appreciated how widely applicable it really is. Such a technique can almost always be reasonably employed provided there exists a reliable formulation of the direct problem. Appendix 6-B invokes the null-field method (see Section 4.3) in order to formulate the direct problem for a totally reflecting object, and solves the inverse problem by the Newton-Kantorovich technique. Appendix 6-C applies the technique to a formulation of the penetrable scattering problem that expands the emanations in basis functions which explicitly incorporate the inverse scattering data.

Techniques that do not strictly fall within the terms of reference of this thesis but are important in theory and in practice are summarised in Section 6.8.

6.2 INVERSE SOURCE PROBLEMS AND THE DIMENSIONALITY DIFFICULTY

The conventional volume-source formulation (see Section 4.2.1) of scattering expresses the scattered field in terms of radiation from an equivalent source density Σ that is proportional to the inhomogeneity (χ^2-1) and to the field ψ within the inhomogeneous region Υ_s,

$$\psi_s(\vec{x},k) = \int_{\Upsilon_s} \Sigma(\vec{x}_1,k)g(\vec{x},\vec{x}_1,k)\,d\Upsilon(\vec{x}_1), \tag{6.1}$$

$$\Sigma(\vec{x},k) = k^2(\chi^2(\vec{x},k)-1)\psi(\vec{x},k). \tag{6.2}$$

The equivalent source density is seen to depend on ψ, which is a general function of position and frequency that itself depends on χ and the incident field as generated by sources under the experimenter's control. The data for inverse source and inverse scattering problems, as defined in Section 1.3, consist of measurements of the scattered field at points comprising a set Υ_m removed from Υ_s. Since the field within Υ_s cannot be probed directly, by definition, Σ depends on the product of two unknown terms, (χ^2-1) and ψ. It is convenient, at least until an estimate for Σ has been obtained, to

forget that ψ_s is a scattered field and to think of it purely as the emanations generated by Σ. That is, one first attempts to ascertain an equivalent source density Σ, whose decomposition into $(\chi^2 - 1)$ and ψ is to be attempted at a later stage [Hoenders 1978]. By this procedure, the inverse scattering problem, in which a single incident field is employed, assumes the form of an inverse source problem. It should be noted that because monochromatic wave motion is being examined, the equivalent source density can be expected to exhibit an arbitrary degree of spatial coherence. It is worth recalling that there is a wide class of important inverse source problems, not directly considered in this thesis because they do not fall within its scope, that consider random, partially coherent [Born and Wolf 1970, Chapter 10], [Beran and Parrent, Jr. 1974], [Carter and Wolf 1985], [LaHaie 1986], and spatially incoherent sources [Bates 1982].

To obtain all the information about Σ available in its radiated field, it is in principle sufficient for Υ_m to be a surface enclosing Υ_s, from which the field throughout the remainder of free-space can then be completely determined [Bates 1984,Wolf and Porter 1986]. This is obvious from Huygens' principle [Baker and Copson 1950] and plays an essential role in generalised holography [Porter 1970,Porter and Devaney 1982a,Porter and Devaney 1982b]. In practice, the presence of noise implies that taking more measurements increases the accuracy of the data, but this is merely reducing additional uncertainty over and above the ideal situation. The presence of evanescent waves [Sherman 1969], [Wolf and Nieto-Vesperinas 1985] means that, in practice, only the homogeneous waves can be stably back-propagated.

To see why it is sufficient for Υ_m to be a surface, consider the expression for ψ_s in any of the separable coordinate systems (see Section 3.3.1). Outside Υ_s then, ψ_s can be written in the form

$$\psi_s(\xi_1, \xi_2, \xi_3, k) = \sum_m A_m(k) y_m^{(o)}(k\xi_1) Y_m(\xi_2, \xi_3). \tag{6.3}$$

Measurement of ψ_s on a surface, say $\xi_1 = R$, is sufficient to determine $\{A_m\}$ by invocation of the orthogonality relation (3.31),

$$A_m = \iint \psi_s(R, \xi_2, \xi_3, k) Y_m^*(\xi_2, \xi_3) w(\xi_2, \xi_3)\, d\xi_2\, d\xi_3. \tag{6.4}$$

Even if Σ is merely a function of ξ_1 $(0 \leq \xi_1 < R)$, so that the expansion (3.32) permits the ξ_2 and ξ_3 integrals to be performed immediately, (6.1) reduces to [Devaney and Sherman 1982]

$$\int_0^R \Sigma(\xi_1', k) y^{(r)}(k\xi_1')\, d\xi_1' = A_m(k). \tag{6.5}$$

Clearly, knowledge of A_m alone is insufficient to determine Σ, implying that the inverse source problem is non-unique. It is not even possible to assign Σ a unique constant value because the value of R can be adjusted so as to produce any desired value for A_m in (6.5). The non-uniqueness arises because an arbitrary Σ is three-dimensional while the measurement surface is effectively only two-dimensional. This discrepancy is referred to by Bates [1984] as the dimensionality difficulty.

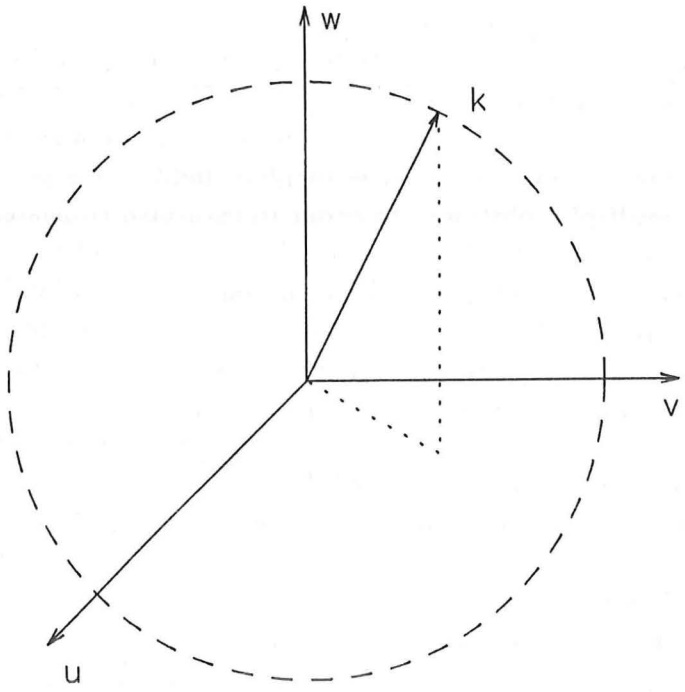

Figure 6.1: The Fourier transform of Σ is a general three-dimensional quantity. Only the Fourier components on the sphere of radius k contribute to the scattered field. The other Fourier components are non-radiating sources.

Another interpretation [Wolf and Porter 1986] relates the quantity,

$$\Gamma(\vec{x},k) = \int_{\Upsilon_m} \left\{ \psi^*(R,\xi_2,\xi_3,k) \frac{\partial g}{\partial n}(\vec{x},R,\xi_2,\xi_3,k) - g(\vec{x},R,\xi_2,\xi_3,k) \frac{\partial \psi^*}{\partial n}(R,\xi_2,\xi_3,k) \right\} d\xi_2\, d\xi_3, \tag{6.6}$$

to the components of the Fourier transform of Σ that lie on the sphere of radius k in Fourier space (see Figure 6.1). The result entails arguments similar to those presented when arriving at (4.8). The said Fourier components of Σ are called the radiating sources while the remaining Fourier components are non-radiating sources [Devaney and Wolf 1973,Kim and Wolf 1986].

Note that Γ is defined in terms of known (to the experimenter) quantities. Thus, knowledge of the scattered field on a surface enclosing Υ_s is not sufficient to enable unambiguous determination of Σ [Bleistein and Cohen 1977], [Devaney and Sherman 1982], [Fischer and Langenberg 1984]. It is worth noting that uniqueness is ensured if the source density is spatially incoherent or composed of discrete scatterers [Bates *et al.* 1985] or if the sources are known to lie on a known surface

(so that the dimensionality difficulty is rectified), but the constraint of finite extent is not sufficient [Devaney and Sherman 1982].

Despite the non-uniqueness of the solution for the inverse source problem, it is often possible to remove the ambiguity for the inverse scattering problem, because there is an additional constraint that the equivalent source can be decomposed into a product of $(\chi^2 - 1)$ and ψ. The decomposition is only consistent if ψ is the field generated by the interaction of χ and the incident field. This permits elimination of some, perhaps all, multiple solutions. In order to establish consistency, it is necessary to exploit another degree of freedom such as the frequency of the incident field or its direction of incidence. Although ψ is different for each incident field employed to probe the inhomogeneous region, then the respective decompositions must each possess the same χ. It is on this extra consistency that the possibility of obtaining unique solutions to inverse scattering problem rests. Note, however, that to verify consistency, it is necessary to determine the fields generated by the interaction of χ and each incident field. This is by no means a trivial task. While this is in no sense a proof of uniqueness, it provides a general indication of the computational approach needed to solve inverse scattering problems, and demonstrates the necessity of possessing reliable methods of solving the direct problem.

In practical applications the computational demands of establishing consistency have generally been deemed too daunting, so that a variety of approximations and heuristic approaches have been adopted in order to make inverse scattering problems tractable. Typically, this involves simplifying the decomposition of the equivalent source density to such an extent that the solution can be inferred without resorting to numerical computation of the total field (see, for example, Ramm [1986]). For example, the first Born and Rytov approximations (see Sections 4.2.1 and 4.2.2 respectively) employ an approximate equivalent source density that is proportional to the product of the inhomogeneity $(\chi^2 - 1)$ and the incident field, the latter being of course known *a priori*. A second example is the use of the projection model, described in Section 5.4, which also removes all the ambiguity in the determination of the equivalent source density.

6.3 REMOTE PROBING COMPUTED TOMOGRAPHY (CT)

Conventional or remote probing CT is the name given to the species of inverse problem that arises when the emanations are actively used by the experimenter to probe the body and the behaviour of the emanations is assumed to be adequately described by propagation along known rays [Brooks and Di Chiro 1976b], [Dines and Lytle 1979a], [Kak 1979], [Greenleaf 1983], or equivalently by the projection model of Section 5.4 [Cormack 1963], [Cormack 1964], [Lewitt and Bates 1978a].

The inverse Radon transform has already been stated (5.24). Another form of the inverse Radon transform that can be derived by manipulation of Fourier transform identities [Bracewell and Riddle 1967], but also from mathematical considerations [Nievergelt 1986], is

$$f(r;\theta) = \mathcal{B}\left\{\tilde{p}(\xi;\phi)\right\}(r;\theta), \tag{6.7}$$

where $\tilde{p}$ is known as the filtered or modified projection,

$$\tilde{p}(\xi;\phi) = \mathcal{F}_1^{-1}\left[|\rho|\mathcal{F}_1[p](\rho;\phi)\right](\xi;\phi), \tag{6.8}$$

the Fourier transforms being one-dimensional, and $\mathcal{B}$ is the back-projection operator,

$$\mathcal{B}\{\tilde{p}(\xi;\phi)\}(\vec{x}) = \int_0^{\pi} \tilde{p}(\xi - \vec{x}\cdot\hat{\xi};\phi)\,d\phi. \tag{6.9}$$

The operation of back-projection, has the effect of smearing, over a ray, the value of the projection corresponding to that ray (see Figure 5.2).

Various other forms of the inverse Radon transform exist [Barrett 1984] but do not lead to the most efficient numerical implementations. The computational schemes based on (5.24), known as direct Fourier reconstruction [Stark *et al.* 1981,Natterer 1985], are generally discarded in favour of more popular schemes based on (6.7), known as filtered or modified back-projection techniques [Lewitt *et al.* 1978,Rowland 1979,Lewitt 1983]. Filtered back-projection has many advantages for software and hardware implementation, notably requiring only one-dimensional Fourier transforms of data that can usually be obtained on regular linear grids, and the utilisation of the back-projection operation.

ART [Gordon *et al.* 1970,Gordon 1974] and SIRT [Gilbert 1972,Norton 1985], [Andersen and Kak 1984] are iterative techniques to solve (5.22) for f, considered as a set of linear simultaneous equations. Other things being equal, filtered back-propagation is generally more computationally attractive. The significance of iterative techniques is that they are amenable to modifications that allow for deviations from the ideal projection model, such as those mentioned in Section 5.4, to be incorporated [Lytle and Dines 1980], [Datta and Bandyopadhyay 1985], [Andersen 1987]

6.4 REMOTE SENSING COMPUTED TOMOGRAPHY

The success of conventional CT (Sections 5.4 and 6.3) to model the transmission of X-rays through biological tissue has lead to attempts to apply similar theory to other species of emanations. Two medical imaging modalities, PET (positron, or paired photon, emission tomography) [Derenzo 1977], [Derenzo *et al.* 1977], [Thompson *et al.* 1979] and SPECT (single photon emission computed tomography) [Knoll 1983], [Larsson and Israelsson 1982], [Muehllehner and Colsher 1981], employ radionuclides that are deposited in the body and whose photon emissions are detected. The distribution of photon emitters acts as a spatially incoherent source distribution so that PET and SPECT are inverse source problems, as defined in Section 1.3. The notion of radiation travelling along straight rays is retained.

The radiation process by which photons are emitted by the body can be described on the microscopic scale by simple mechanisms [Stanton 1969, Chapter 4]. In PET, positrons (particles of equal mass but opposite charge to electrons) are emitted to be annihilated by electrons after travelling a few millimetres. Energy and momentum are conserved by the creation, for each annihilation event, of two photons of equal energy and almost mutually opposite directions. Coincidence circuitry enables inference of the ray traversed by the two photons [Budinger *et al.* 1977], [Thompson *et al.* 1986].

In single photon emission CT (SPECT), the radionuclides emit individual photons [Budinger and Gullberg 1977]. Transition to the macroscopic scale is far from understood, often only being amenable to statistical description [Floyd *et al.* 1984], and many approximations must be invoked in order to pose a tractable problem.

To draw parallels between the projection arising in conventional CT and measurements that can be realised in SPECT, it is necessary to introduce collimators that confine, to a pencil of rays, the region contributing to detected radiation [Rotenberg and Johns 1965]. An acceptable level of detection is only achieved when the beamwidth of the collimators is appreciably larger than that of X-ray collimators [Derenzo and Budinger 1977]. Compton scatter is a significant contribution to the detected emanations and tends to further increase the effective beamwidth of each detector [Floyd Jr. *et al.* 1985]. As in conventional CT, the emanations suffer attenuation during passage through the body. It is assumed that the spatial distribution of the attenuation coefficient has been previously determined by the conventional CT schemes summarised in Section 6.3. Correction for attenuation is an area of current research interest. The majority of research into SPECT assumes a constant attenuation coefficient. An original contribution to the relation between the detected emanations and the source distribution that simultaneously takes into account the finite and variable resolution of the detectors and a constant attenuation coefficient [Tan *et al.* 1987] is reproduced as Appendix 6B. This is significant in that previous approaches consider constant attenuation and infinitesimal beamwidth [Tretiak and Metz 1980], [Clough and Barrett 1983], which permits approximate attenuation correction to be performed by a variety of adaptations of the conventional CT techniques [Hsieh and Wee 1976], [Chang 1978], [Lewis *et al.* 1982], [Tanaka *et al.* 1984], [Zeeberg *et al.* 1985], [Axelsson *et al.* 1987].

6.5 DIFFRACTION TOMOGRAPHY (FOURIER IMAGING)

Fourier imaging [Robinson 1982] is the name given to the species of inverse problem that arises when the direct problem is adequately described by the Fourier transform relation (4.12) derived within the first Born approximation. Recall that a similar expression is obtained for the first Rytov approximation (refer to Section 4.2.2). Because the model employed explicitly accounts for diffraction (but, nevertheless, only approximately) and can be regarded as the finite wavelength limit of computed tomography, Fourier imaging has also been called diffraction tomography [Mueller 1980], [Devaney 1982], [Soumekh 1986a]. Both the (first) Born and Rytov approximations relate (a functional of) the scattered field to the generalised constitutive parameter χ by a (linear) Fourier transform identity of the form (4.12). When two spatial dimensions are considered, (4.12) relates the one-dimensional Fourier transform of the scattered field to the two-dimensional Fourier transform of χ on a circle in Fourier space, and is often termed the diffraction-slice theorem [Devaney 1982,Devaney 1985a], in analogous fashion to the projection (projection-slice) theorem (5.23).

The Fourier transform of the scattered field provides the Fourier transform of χ on circles (spheres) in Fourier space, rather than on lines (planes). At a fixed frequency, maximal information concerning the Fourier transform of χ is gathered by

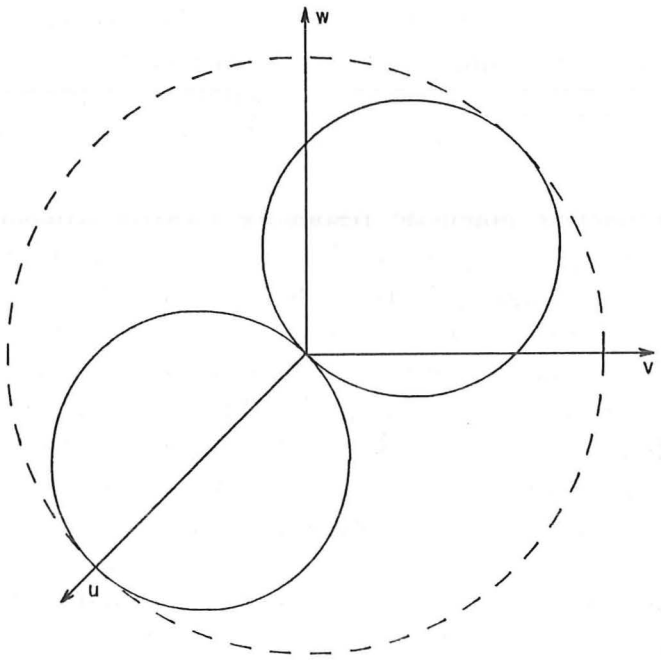

Figure 6.2: The scattered field from one incident field provides Fourier components of χ situated on a sphere. By varying the incident direction, Fourier space is covered within the Ewald limiting sphere, allowing an inverse Fourier transform to infer χ.

employing incident waves of all angles, as in conventional CT (see Figure 6.2).

For finite k, however, coverage of Fourier space extends only to the sphere in Fourier space of radius $2/\lambda$, known as the Ewald limiting sphere [Wolf 1969], because for real angles, the magnitudes of sines and cosines cannot exceed unity. This is in keeping with the intuitive notion that the spatial variations of the medium's constitutive parameters significantly smaller than a wavelength cannot be discerned. The inverse Fourier transform of (4.12) is a band-limited approximation to the true distribution of χ. Comparison of (4.28) with (5.25) reveals that conventional CT can be considered as the high frequency limit of diffraction tomography [Devaney 1983] where circles of zero curvature become straight lines.

Numerical techniques based on direct Fourier inversion of (4.12) suffer from problems, similar to direct methods in conventional CT, of interpolating onto a rectangular grid [Pan and Kak 1983]. Devaney [1982] has pioneered the back-propagation algorithm which is analogous to filtered back-projection. On account of the diffraction effects implicit in Fourier imaging and diffraction tomography models, it is necessary to employ filters whose characteristics vary with distance from where the field is back-propagated. Applications with various measurement geometries are have also attracted interest [Devaney and Beylkin 1984,Lan *et al.* 1985,Baribaud *et al.* 1985,Paoloni 1987]. Diffraction tomography has recently been examined in the context of vector wave mo-

tion [James *et al.* 1985,James 1986] and extended to an iterative procedure [Devaney 1986].

Some of the more apparent deficiencies in the diffraction tomography model, including the lack of explicit accommodation of multiple scattering and attenuation, have been investigated by applying reconstruction algorithms to scattering data obtained for simple geometries [Azimi and Kak 1983,Slaney *et al.* 1984,Soumekh and Kaveh 1986], [Paoloni 1986].

Imaging by nuclear magnetic resonance (NMR) [Hinshaw and Lent 1983], [Straughan 1986] or magnetic resonance imaging (MRI) as it is now often called [Williams 1984], can be arranged [Morris 1986, Chapter 2] to provide data which represent samples, of the Fourier transform of the density, on a rectangular grid whose spacings, in three orthogonal directions, satisfy the sampling theorem (see Section 2.7). This means that MRI can be viewed as a species of Fourier imaging. It is worth keeping this in mind in studying the MRI literature, although most MRI image reconstruction algorithms operate on differently organised data sets [Cho *et al.* 1987].

The far scattered field from a totally reflecting body, computed on the basis of planar physical optics (see Section 4.4), can be expressed as a Fourier integral over the equivalent source density, provided one is only interested in the scattering within some 60 degrees of the major peak of the scattered field, so that the source density can be inferred by Fourier transformation [Lewis 1969]. Hence, physical optics far field inverse scattering can be regarded as a species of Fourier imaging. The Fourier integral representation can remain valid for bodies that are only partially reflecting [Beckmann and Spizzichino 1963, Chapter 3]. Bates and Wall [1977b] review this in greater detail.

Approximate inversion algorithms, valid for fields spanning all angles, can be constructed on the basis of the extensions to physical optics discussed in Section 4.4. While they are less computationally convenient than the planar physical optics algorithms, because they cannot be formulated as integral transforms, Bates and Wall [1977c] have show that stable inversion is practicable, at least for circular physical optics.

6.6 SPECKLE IMAGING

It is often the case in imaging that the object of interest is a part of the generalised constitutive parameter consisting of discrete scattering centres embedded in an inhomogeneous refractive index. When only the juxtaposition of the scatterers and an indication of their relative amplitudes is required, it is often convenient to consider the inhomogeneous medium merely as a source of distortion. The distortion is often speckly in nature. The information about the nature of the medium conveyed by speckle is often disregarded. Rather, speckle is treated as unwanted artefact to be removed from the image. The problem of dealing optimally with speckle has come to be called "speckle processing" or "speckle imaging" [Bates 1982].

Speckle has the appearance of a random phenomenon. Techniques to reduce speckle artefacts generally require an ensemble of images, whose distortions are statistically independent. These images are manipulated in such a manner that spurious features are averaged out while the desired features are preserved. The de-

tails of such techniques are numerous and varied, and depend largely on the peculiarities of the application under consideration, primarily the origin of the speckle and its statistical behaviour depend on the precise mechanism causing the speckle [Wells and Halliwell 1981,Robinson and Knight 1981,Melton Jr. and Magnin 1984], [Ward and Saleh 1985]. The remainder of this section concentrates on the aspect most germane to the inverse scattering techniques of the previous chapter, namely, approaches that reduce distortion caused by shortcomings of techniques based on echo-location.

The shift-and-add (SAA) technique described in Section 5.2 operates on an ensemble of images formed by some or other imaging procedure and is a form of supplementary processing that can reduce distortion caused by imperfections and limitations of the procedure employed. When viewing an object through a fluctuating medium, the ensemble of distorted images on which SAA operates can be recorded sequentially, without altering the band of frequencies (or, equivalently, wavelengths) accepted by the detection apparatus. In a particularly promising application of SAA, wide-angle ultrasonic scatter imaging [Robinson and Bates 1980], it is impractical to gather a temporal sequence of independently distorted images. It often transpires, however, that images formed within different frequency bands are adequately independent [Bates and Robinson 1981]. The version of the true image implicit in each of the blurred images is different, of course, because each is formed within a different band of frequencies. For the remainder of this section, the dependence of the fields describing the emanations at the m-th frequency is denoted by a subscript. The major difference is in phase$\{f(\vec{x})\}$. The phase differences can be compensated in the simple manner described below.

Consider the geometry of Figure 6.3. The scattering region Υ_s consists of a collection of point scatterers embedded in an inhomogeneous medium,

$$k^2\chi^2(\vec{x}) = \sum_{n=1}^{N} f_n \delta(\vec{x} - \vec{x}_n) + \nu^2. \tag{6.10}$$

An arbitrary point within the measurement region Υ_m, which is taken to enclose Υ_s, is denoted in polar coordinates by $(R;\Theta)$. The use of a subscript in "Υ_m" does not signify that the measurement region varies with each speckle image. The incident field is a plane wave whose direction of propagation defines the η-axis,

$$\psi_i(\vec{x}, k) = e^{-ik\eta}. \tag{6.11}$$

The Cartesian coordinates (ξ, η) are a rotation, through the origin and by an angle ϕ, of the (x, y) system. The m-th speckle image is generated by forming, on a family of surfaces, the equivalent source distribution that produces the m-th measured field. In two spatial dimensions, these surfaces are taken to be lines of constant ξ. When R is a constant large enough for Υ_m to be in the far-field of Υ_s, so that (3.27) applies, and assuming that the scattering region is small enough to permit the approximation

$$|\vec{x} - \vec{x}_n| \approx R - \vec{x}_n \cdot \hat{x}, \tag{6.12}$$

to be invoked whenever $\vec{x} \in \Upsilon_m$ and $\vec{x}_n \in \Upsilon_s$, then the normalised radiations $\Psi_m(R;\Theta)$ received due to an equivalent source distribution $e^{-ik_m\eta}s_m(\xi_1, \eta)$ are given by

$$\Psi_m(R;\Theta) = e^{ik_m\xi_1\cos(\Theta-\phi)} \int_{-\infty}^{\infty} s_m(\xi_1, \eta) e^{-ik_m\eta[1-\sin(\Theta-\phi)]}\, d\eta. \tag{6.13}$$

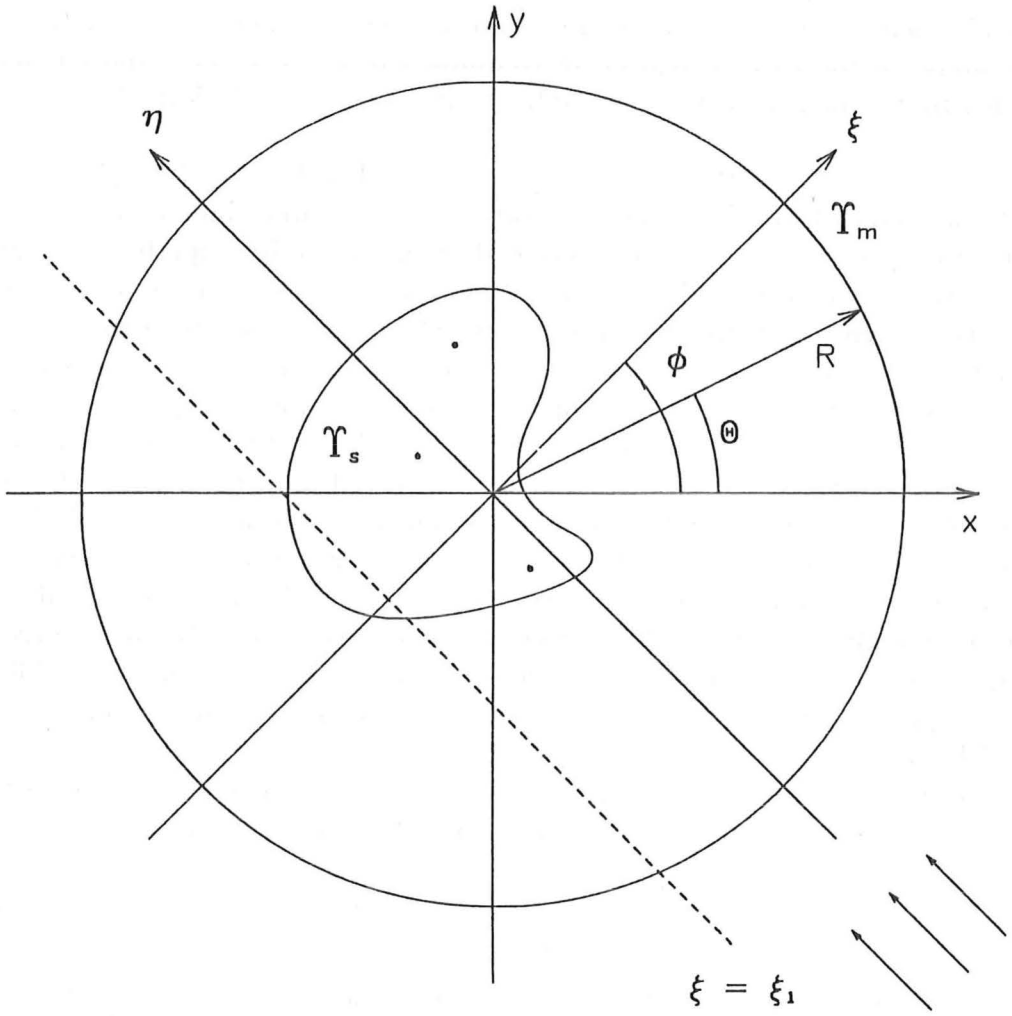

Figure 6.3: Geometry for coherent speckle imaging. The scattering region Υ_s consists of a collection of point scatterers embedded in an inhomogeneous medium. The measurement region Υ_m is a circle enclosing Υ_s. The incident field is a plane wave travelling in the positive η-direction.

The infinite limits on the integral merely imply that the whole of Υ_s along the line $\xi = \xi_1$, which is arbitrary but fixed, is included. On associating Ψ_m with the m-th measurement of the scattered field, and introducing

$$u = k_m[1 - \sin(\Theta - \phi)], \tag{6.14}$$

$$F_m(u) = \Psi_m(R; \Theta), \tag{6.15}$$

inversion of (6.13) is easily found to be

$$s_m(\xi_1, \eta) = \int_{-\infty}^{\infty} e^{i\xi_1\sqrt{1-(k_m-u)^2}} F_m(u) e^{iu\eta}\, du, \tag{6.16}$$

which is in the form of a Fourier transform relation. The choice of $e^{ik\eta}s_m(\vec{x})$ for the equivalent source density ensures that, when (i) the scatterers fall on a line $\xi = \xi_1$, (ii) ν is identically unity, and (iii) the first Born approximation applies, $s_m(\xi_1, \eta)$ is in perfect agreement with the first term of (6.10).

The presence of distortion implies that, in practice, the measured quantity is conveniently expressed as

$$S_m(u) = F(u)H_m(\vec{u}) + C_m(\vec{u}), \tag{6.17}$$

The term $H_m(\vec{u})$ represents the distortion experienced similarly by all radiation arriving at $\vec{u}$, and is such that S_m is linear with respect to $\Psi(\vec{u}, k_m)$, whereas $C_m(\vec{u})$ includes all other imperfections, such as random noise, nonlinear effects, and errors introduced by model approximations. In particular, H_m is taken to contain a description of the distortion experienced by the first order scattering while C_m describes higher order multiple scattering. The inverse Fourier transform of (6.17) results in an expression of the form (5.6).

The motivation behind incorporating multiple scattering terms in C_m is chiefly because they can be expected to exhibit a large amount of destructive interference when averaged over independent speckle images. The field at $\vec{x}_m$ can be written as [Ishimaru 1978, Chapter 14]

$$\psi_s(\vec{x}_n, k_m) = \sum_{\substack{p=1 \\ p \neq n}}^{N} e^{ik_m r_{np}} + \text{higher order terms}, \tag{6.18}$$

where $r_{np} = |\vec{x}_n - \vec{x}_p|$. Despite the normalisation that accounts for the phase factor $(-ik\eta)$ introduced by the incident field, the terms in (6.18) exhibit phase variation that depends on k_m. Actual experiments performed [Bates and Minard 1984] obtain the ensemble of images from the use of a single incident direction, $\phi = 0$, and demonstrate the practical efficacy of the technique.

Suggestions for future investigation of speckle processing are made in Section 7.2.

6.7 ITERATIVE ARBITRARILY ACCURATE SOLUTIONS

When it is found impossible to form an explicit expression for the inverse (5.1) of the direct mapping, the only recourse is to attempt iterative solution. By treating (4.1) as a

nonlinear equation in an abstract normed linear space, the full power of sophisticated mathematical techniques can be invoked to derive iterative algorithms that lead to convergent sequences of estimates of the true generalised constitutive parameter.

A common form of iterative algorithm is the fixed-point algorithm [Rall 1969, Chapter 2]. The direct mapping (4.1) is first rewritten in the form

$$\chi = \mathcal{H}(\chi). \tag{6.19}$$

Given an initial estimate $\chi^{(0)}$, the fixed-point iterates are defined for $l > 0$ by

$$\chi^{(l)} = \mathcal{H}(\chi^{(l-1)}). \tag{6.20}$$

The limit point of the iterates, if they in fact converge, is known as the fixed-point of $\mathcal{H}$ and is a solution of the original problem (4.1). The general direct mapping is nonlinear with respect to χ and a form for $\mathcal{H}$ that ensures convergence of the fixed-point algorithm may be hard to establish.

The Newton-Kantorovich algorithm [Rall 1969, Chapter 4] is an example of the fixed-point technique to solve nonlinear operator equations whose most familiar manifestation is, perhaps, Newton's solution of nonlinear algebraic equations. Under quite wide conditions, the algorithm can be shown to converge quadratically. It can also be applied equally satisfactorily to regularised forms of ill-posed problems. Consider a general nonlinear operator equation

$$\Omega(x) = \mathcal{D}. \tag{6.21}$$

The data for an inverse problem are represented by $\mathcal{D}$. Given a description of the direct mapping to define the operator Ω, the inverse problem is posed as solving (6.21) for x, which belongs to a normed linear space. At each iteration, an improved estimate of the solution is obtained by setting

$$x^{(l+1)} = x^{(l)} + \lambda^{(l)} \delta x^{(l)}. \tag{6.22}$$

The effect of a perturbation $\delta x^{(l)}$ to $x^{(l)}$ is approximated by linearising Ω about the current estimate $x^{(l)}$, i.e.

$$\Omega(x^{(l)} + \delta x^{(l)}) \approx \Omega(x^{(l)}) + \Omega'(x^{(l)}) \delta x^{(l)}, \tag{6.23}$$

where Ω' denotes the Fréchet derivative of Ω [Rall 1969, Chapter 3]. To compute an improved estimate, (6.23) is equated to the right hand side of (6.21), and the resulting equation

$$\Omega'(x^{(l)}) \delta x^{(l)} = \mathcal{D} - \Omega(x^{(l)}), \tag{6.24}$$

is solved (in a least squares sense if over-determined) for the update $\delta x^{(l)}$. For the conventional Newton-Kantorovich algorithm, $\lambda^{(l)}$ is always unity. Improved convergence to a solution may be obtained by performing a line search to determine an optimal $\lambda^{(l)}$ at each iteration. Techniques such as the Levenberg-Marquardt algorithm [Dennis and Schnabel 1983] may also improve the algorithm. An important factor in the success of the basic method is the faithfulness of the initial approximation $x^{(0)}$.

Appendices 6-B and 6-C represent useful extensions to the Newton-Kantorovich inverse scattering literature in that they demonstrate, respectively, the technique can

actually be implemented for the Bates and Wall [1977a] formulation of the null-field method and for what is perhaps the most general implicit solution [Bates 1975b], [Bates 1984] to the inverse scattering problem for macroscopic wave motion. Related applications of the Newton-Kantorovich algorithm to the inverse problem considered in Appendix 6-B are due to Roger [1981] and Kristensson and Vogel [1986] . Connolly [1985] and Connolly *et al* [1985] review all this in considerable detail.

6.8 OTHER TECHNIQUES

Whilst pertinent references are included in previous chapters, there are many other ingenious techniques reported that cannot all be covered here. The continued prominence, and many of the trends, of inverse theory can be appreciated by perusal of the special issues of journals, often concerning diverse scientific and technological applications, that are devoted to aspects of inverse theory. Some of the more notable ones are :

- Acoustic Imaging: [Wang 1979]
- Applications of Electromagnetic Theory to Geophysical Exploration: [Wait 1979]
- Computerized Tomography: [Herman 1983]
- Digital Acoustical Imaging: [Wade *et al.* 1984]
- Direct Problems in Propagation and Scattering: [Mittra 1985]
- Electrical Impedance Tomography — Applied Potential Tomography: [Brown *et al.* 1987]
- Electromagnetics in Applied Geophysics: [Howard Jr. 1984]
- 1983 International Geoscience and Remote Sensing Symposium (IGARSS 1982): Remote Sensing: The promise of Remote Sensing: [Sieber 1983]
- 1984 International Geoscience and Remote Sensing Symposium (IGARSS 1983): Remote Sensing: Extending Man's Horizon: [Rea 1984]
- Physics and Engineering in Nuclear Magnetic Resonance Imaging: [Cho and Nalcioglu 1984]
- Inverse Methods in Electromagnetics: [Boerner *et al.* 1981]
- Inverse Problems in Propagation and Scattering: [Devaney 1985b]
- Medical Imaging: [Wells 1987,Dallas and Wagner 1987]
- Medical Ultrasound: [Waag 1983]
- Non-destructive Testing: [Bond and Reynolds 1987]
- Ocean Acoustic Remote Sensing: [Ehrenberg 1986]

- Physical and Computational Aspects of 3-Dimensional Image Reconstruction: [Cho 1974]
- Radar Reflectivity: [Fritsch 1965]
- Signal Recovery: [Fienup 1983a,Fienup and Rushforth 1987]
- Technology and Health Care: [Sherman 1979]

This remainder of this section lists certain key references relating to aspects of inverse theory which, although of considerable scientific and technological importance, do not come strictly under the purviews of this thesis. They are ordered under the following headings with the minimum of explanation.

6.8.1 Gel'fand-Levitan and related methods

Inverse scattering theory arose originally in a quantum mechanical context (for a chronological survey, see the Foreword to [Chadan and Sabatier 1977]) wherein, as explained in Section 2.4, the refractive index is effectively everywhere unity. The early work was concerned with exact methods of solution for one-dimensional or spherically symmetric situations [Chadan and Sabatier 1977,Newton 1981b]. The celebrated Gel'fand-Levitan and Marchenko techniques have been recently reviewed by Hashaby and Mittra [1987]. Newton [1980,1981c,1982a,1985] and others [Rose *et al.* 1984], [DeFacio and Rose 1985,Rose *et al.* 1986], have recently extended these to three spatial dimensions. The relation between these integral equation techniques, so-called "layer-stripping" techniques (based on a differential equation formulation of the direct problem [Yagle and Levy 1984], [Yagle and Levy 1986]), and also the Radon transform (see Section 5.4), has been emphasised by Yagle [1986].

6.8.2 Reconstruction from resonant frequencies

Kac [1966] once asked the delightful question "Can one hear the shape of a drum?" The implication is that, given the resonant frequencies of an object, one should be able to deduce its constitution and shape. Significant progress seems only to have been achieved for one-dimensional problems [Lytle and Lager 1976], [Barcilon 1974], [Barcilon 1976], [Barcilon 1979], [Barcilon 1982], [Barcilon 1987], [Gladwell 1984], [McLaughlin 1984b], [McLaughlin and Rundell 1987]. This topic has recently been reviewed by McLaughlin [1986].

6.8.3 Conservative fields

At first sight, physical intuition suggests that the effectiveness of any imaging scheme increases with the reduction of the wavelength of the radiation employed. Surprisingly, perhaps, imaging techniques have been devised that employ effectively infinite wavelength [Price 1979], [Liu and Chen 1984], [Barber and Brown 1984], [Kohn and Vogelius 1984], [Kohn and Vogelius 1985], [Connolly 1985], [Seagar and Bates 1985], [Connolly and Wall 1987], although similarities with conventional CT (refer to Sections 5.4 and 6.3) are somewhat illusory [Bates *et al.* 1980],

[Bates *et al.* 1983]. This has been taken advantage of rudimentarily by geophysicists for quite some time [Dines and Lytle 1981], [Wait 1983], [Smith and Vozoff 1984], [Assal and Mahmoud 1987]. Recently, remarkable success has been achieved in imaging the interior of the human body [Seagar 1983], [Brown *et al.* 1985]. The best recent account of clinical applications is [Brown *et al.* 1987].

6.8.4 Existence and Uniqueness Results

Conditions ensuring existence and uniqueness of the solution to the inverse problem for an impenetrable obstacle have been obtained by Schiffer (see Lax and Phillips [1967, Chapter 5]) and extended by Jones [1985]. The related problem of characterising inverse scattering data has been examined by Ramm [1987]. Such results contain theoretical arguments and do not suggest actual algorithms to solve inverse problems, particularly when noise is present in the data.

Difficult though this question is, it must be worth pursuing because, if sufficiently manageable results can be obtained, they can be expected to help one simplify and/or generalise inversion algorithms.

6-A ALLOWING FOR VARIABLE RESOLUTION AND CONSTANT ATTENUATION IN SPECT

Allowing for variable resolution and constant attenuation in SPECT

D.G.H. Tan, BSc
J.X. Qu, ME
K.L. Garden, ME, PhD
Prof. R.H.T. Bates, DSc(Eng), FEng, FRSNZ, FIEE

Indexing terms: Image processing, Biomedical applications, Biological effects, Computer applications, Computed tomography

Abstract: By representing an image as a distribution of Gaussian blobs, due account is taken, for a single photon emission computed tomography (SPECT) system, of finite detector resolution (leading to spatially varying image resolution) and constant attenuation of the radiations on their passage through a body. A simple enhancement procedure is proposed. It is argued that the effects of variable attenuation (as must occur in practice) are unlikely to be important in general.

1 Introduction

Single-photon-emission computed tomography (SPECT) has been investigated for more than a dozen years [1, 2]. From the beginning, there has been interest in compensating for the attenuation suffered by the radiations on their passage through the body. A significant difficulty is that, even if conventional computed tomography (CT) measurements [3] are made initially, using radiation of the same energy as that responsible for the SPECT image, there is no guarantee that a convergent algorithm exists to account for the attenuation [1, 4]. Although the general attenuation problem for SPECT remains unsolved (and may well be unsolvable), the constant attenuation problem is well understood now [5–7]. In this latter problem, one assumes that the attenuation coefficient is invariant throughout the body. This assumption only leads to useful practical results if the surface of the body is known from prior observation; using calipers or by echo location, for instance [4]. Some success has been achieved with post-processing based on assuming different values of the attenuation coefficient in different parts of the body [5].

Received signal levels are usually high enough in conventional CT that one can use collimated detectors possessing narrow beams. It thus makes good physical sense to think of the radiation received by an individual CT detector as being confined to a thin pencil of rays. SPECT signals, on the other hand, tend to be photon-limited, so that it is usually necessary to employ detectors having appreciable beamwidths. Surprisingly, the effect of a finite SPECT beamwidth, which leads to variable resolution within the body [9], seems to have received comparatively little attention [10, 11].

In this paper we examine both constant attenuation and variable resolution. We also discuss the kind of image processing which can ameliorate their deleterious consequences. Of special significance is that we allow simultaneously for both effects without having to resort to iterative procedures. Despite the greater generality of our analysis, we find ourselves able to propose a very simple enhancement strategy which, nevertheless, appears to be potentially useful. The results of this processing could probably be slightly improved by any of the commonly practised post-processing image-restoration schemes [4–13].

The general SPECT problem is formulated in Section 2, while Section 3 lists the simplifications which we find are necessary to obtain quantitative image-processing criteria. Section 4 presents some general analysis from which specific results are obtained in Sections 5 and 6. The general SPECT problem is re-examined in Section 7 where we conclude that it may not be worth worrying too much about trying to take due account of variable attenuation, because little advantage is likely to accrue even if a practicable means of accomplishing it could be devised. We assess, in Section 8, the practical implications of our approach.

2 General SPECT problem

The closed curve C in Fig. 1 represents the perimeter of a particular cross-section of a 3-dimensional body containing a distribution of radiating sources. The distribution is denoted by $f(x, y, z)$, where the Cartesian co-ordinates x, y and z are taken to be fixed in the cross-section. The ξ, η Cartesian co-ordinates are rotated in the plane of the cross-section by an (arbitrary) angle ϕ with respect to the x, y co-ordinates. The Cartesian and polar co-ordinates of the arbitrary point P in the body are (x, y, z) and $(r; \theta, z)$, respectively. Note that a semicolon always precedes an angular variable in this paper, so that we can write

$$f(x, y, z) = f(r; \theta, z) \quad (1)$$

without being in any way ambiguous concerning the types of co-ordinate implied in each case.

Paper 5064A (S9, E4), first received 7th January and in revised form 9th April 1986

The authors are with the Electrical & Electronic Engineering Department, University of Canterbury, Christchurch 1, New Zealand

J.X. Qu is on leave from the Radio & Electronics Department, University of Science & Technology of China, Hofei, Anhwei, People's Republic of China

We postulate a linear array of identical detectors with their faces (i.e. collecting apertures) in a plane which is parallel to the ξ', z-plane. The ξ-axis, which lies in the x, y-plane at a distance R from the ξ'-axis (to which it is parallel), passes through the centres of all detectors in the array. The centre of a typical detector lies at Q. We

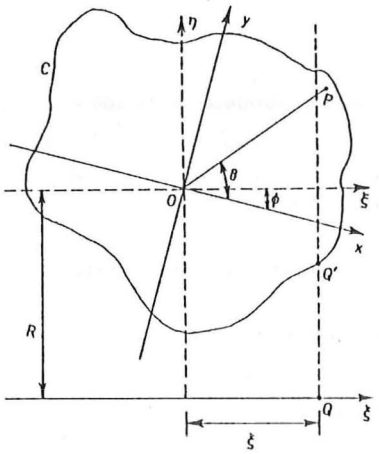

Fig. 1 *Basic geometry and co-ordinate systems for SPECT*
C is the perimeter of a particular cross-section through a 3-dimensional body

denote its total response to the complete distribution of sources within the body by $p_s(\xi; \phi)$. In the absence of any attenuation of the radiations from these sources, $p_s(\xi; \phi)$ can be expressed quite generally as

$$p_s(\xi; \phi) = \int_{-\infty}^{\infty}\int\int f(\xi' \cos(\phi) - \eta \sin(\phi), \xi' \sin(\phi) + \eta \cos(\phi), z) \times D(\xi - \xi', \eta + R, z)\, d\xi'\, d\eta\, dz \quad (2)$$

where the infinite limits here (and in similar instances later in this paper) merely imply that all of the body is being considered, and where $D(\xi - \xi', \eta + R, z)$ is the detector's response to a point source at P. Note that, in the arguments of $f(x, y, z)$ in eqn. 2, the co-ordinates x and y have been expressed in terms of ξ' and η.

We imagine a continuous array of detectors, which can be effectively realised either by scanning a single detector along the ξ-axis (this is of course impracticable in the real medical world because it would take too long) or with a discrete array whose spacing is close enough to satisfy whatever resolution constraints appertain in any particular application (this is practical) [14]. We also postulate that responses are recorded for all ϕ, which is another idealisation which can be effectively realised, in practice, if responses are recorded for discrete values of ϕ, again spaced sufficiently closely to satisfy the prevailing sampling constraints [15].

The general SPECT problem is to recover $f(x, y, z)$ from $p_s(\xi; \phi)$, as defined by eqn. 2, recorded in all planes z for $-\infty < \xi < \infty$ and $0 \leqslant \phi < 2\pi$, on the understanding that the attenuation coefficient $2\pi\,\mu(x, y, z)$, corresponding to each point P in the body, is unknown *a priori*. As we have no idea at present how to attack this problem, and because it probably does not possess a unique solution, it makes sense to pose the restricted SPECT problem, i.e. recover $f(x, y, z)$ when $\mu(x, y, z)$ has been reconstructed previously from a set of conventional CT measurements. It is worth emphasising that the conditions under which even this restricted problem can have a unique solution are as yet unclear. This is why we lower our sights and only give detailed consideration to the simplified problem introduced in Section 3.

Like everyone before us, we have to introduce approximations before we can even begin to analyse the problem. In particular, we note that it is far from clear how to modify the integral in eqn. 2 so as to take account of attenuation. It cannot be done with complete generality by merely inserting a factor such as

$$h(P, Q) = \exp\left(-2\pi \int_{P \to Q} \mu(x, y, z)\, dl\right) \quad (3)$$

into the integrand. This factor represents the attenuation along the straight line path from an arbitrary point in the body to the centre of a typical detector. We are not suggesting that the radiations do not travel in straight lines, just emphasising that the face of a real-world detector has a finite area. Consequently, the general SPECT problem is even more challenging than is customarily intimated [4, 8, 10, 11].

3 Simplified problem

Because the detectors used in SPECT systems have moderate beamwidths (between 5° and 10° say), and as R must be roughly twice the radius of the circle which circumscribes C (to accommodate patients of various shapes and sizes), and owing to the comparatively low resolutions and poor signal-to-noise ratios achieved in practice, we feel it is reasonable to adopt the simplifications introduced here.

Each detector (specifically the one with its centre at Q) is characterised by a simple beam emanating from the detector's centre with its axis parallel to the η-axis. While the width of the beam increases with distance from Q, the shape of the beam is circularly symmetric in any plane parallel to the z, ξ-plane. This means that we now can, in fact, account for attenuation by inserting into the integrand of eqn. 2 the factor $h(P, Q)$ defined by eqn. 3. Note that dl in eqn. 3 is the element of length along the straight line path from P to Q.

The most important of the physical processes affecting the quality of SPECT imagery is Compton scattering. It gives rise to the major part of the attenuation suffered by the radiations on their passage through the body. It also affects the effective beamwidths and beam shapes of the detectors owing to multiple scattering. A ray proceeding from a particular source towards a particular detector may be scattered out of its initial path and then be rescattered into another detector. This increases the effective width of each detector's beam [12]. As the distribution of Compton-scattering material, which is approximately proportional to $\mu(x, y, z)$, varies throughout the body, the beamwidth and beam shape of each detector must, in general, be different, and must alter as ϕ changes. By the central limit theorem, however, the average beam shape can be expected to be Gaussian. As the beam shape in the absence of multiple scattering is close to Gaussian [16], it is reasonable to assume the overall average beam shape to be Gaussian, which is what we do in the following.

The distance from any point within the beam to Q is approximated by R plus the η co-ordinate of the point. This means that the beam can be expressed in the general form $\Phi([(\xi' - \xi)^2 + z^2]/(\eta + R)^2\theta^2)$, where θ is the effec-

tive beamwidth. Radiation emanating from a point source spreads out spherically, weakening according to the inverse square of distance from the source, so that the detector's response to a point source is given by

$$D(\xi - \xi', \eta + R, z) = (1/4\pi(\eta + R)^2)\Phi([(\xi' - \xi)^2 + z^2]/(\eta + R)^2\theta^2) \quad (4)$$

It is also assumed that the straight-line path implied by the integral in eqn. 3 can be taken parallel to the η-axis, so that dl is replaced by $d\eta$, i.e.

$$h(P, Q) = \exp\left(-2\pi \int_{-R}^{\eta} \mu(x, y, z)\, d\eta\right) \quad (5)$$

The distribution $f(x, y, z)$ is assumed independent of z over the beam, so that

$$f(x, y, z) = f(x, y) \quad (6)$$

implying that the distribution of sources throughout a 3-dimensional body can be built up from 2-dimensional distributions existing in stacked parallel cross-sections, each of which can be reconstructed from a separate set of measurements, as in conventional CT [3].

The attenuation is taken to be zero outside the body and constant inside (as already intimated in Section 1, this simplification has to be made in order for us to make any headway). The constant attenuation coefficient is $2\pi\bar{\mu}$. The curve C is understood to be determined by prior measurement, implying that the position of the point Q' (see Fig. 1) is known. The upshot is that $(-R)$ and $\mu(x, y, z)$ are replaced in eqn. 5 by $\bar{\eta}(\xi; \phi)$ and $\bar{\mu}$, respectively, where $\bar{\eta}(\xi; \phi)$ is the η co-ordinate of Q'. Furthermore, $h(P, Q)$ can be normalised by multiplying it by $\exp(-2\pi\bar{\mu}\bar{\eta}(\xi; \phi))$, thereby giving

$$h(P, Q) = \exp(-2\pi\bar{\mu}\eta) \quad (7)$$

because η is, by definition, the η co-ordinate of P.

We have now reached the same position with regard to attenuation as those [5, 7, 17, 18] who have gone before us, but we are simultaneously taking account of the finite beamwidth of the detectors. It is impracticable to attempt to recover the form of $f(x, y)$ from $p_s(\xi; \phi)$ until a functional form for $\Phi(\cdot)$ is specified. For the reasons given in the third paragraph of this section, and because we are approximating $\mu(x, y, z)$ by $\bar{\mu}$, we feel it is appropriate to take each detector to be an average detector, which is characterised by a Gaussian beam shape having an effective beamwidth denoted here by θ. On this understanding, $p_s(\xi; \phi)$ is seen to play the part, familiar in conventional CT [3], of the projection at angle ϕ. Accordingly, we call it the SPECT projection which, by incorporating eqns. 4 to 6 into eqn. 2, can be written as

$$p_s(\xi; \phi) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int f(\xi' \cos(\phi) - \eta \sin(\phi), \xi' \sin(\phi) + \eta \cos(\phi)) \exp\{-2\pi\bar{\mu}\eta - [(\xi' - \xi)^2 + z^2]/2(\eta + R)^2\theta^2\}\, d\xi'\, d\eta\, dz/4\pi(\eta + R)^2 \quad (8)$$

The simplified SPECT problem is to recover $f(x, y)$ from $p_s(\xi; \phi)$, as defined by eqn. 8, given for $-\infty < \xi < \infty$ and $0 \leqslant \phi \leqslant 2\pi$, for particular values of $\bar{\mu}$, R and θ.

Whenever we refer in the rest of the paper to the actual reconstruction of images, we find it convenient (for ease of exposition) to do so in terms of direct Fourier inversion [15]. We emphasise, however, that the more efficient modified back-projection approach remains applicable, in any of the manifestations which are now standard for conventional CT [1, 9].

4 Analytical considerations

Note that the z-integration in eqn. 8 can be done immediately. To proceed further we introduce, in the spirit of the projection theorem of conventional CT [3, 15], the 1-dimensional Fourier transform, with respect to ξ, of $p_s(\xi; \phi)$. We denote this transform by $F_s(\rho; \phi)$. Adopting $\exp(j2\pi\rho\xi)$ for the Fourier kernel leads to

$$F_s(\rho; \phi) = (\theta^2/2) \int_{-\infty}^{\infty}\int f(\xi' \cos(\phi) - \eta \sin(\phi), \xi' \sin(\phi) + \eta \cos(\phi)) \exp(2\pi[j\rho\xi' - \bar{\mu}\eta - \pi(\eta + R)^2\theta^2\rho^2])\, d\xi'\, d\eta \quad (9)$$

It is now instructive to state the result of a particular rearrangement of the integral in eqn. 9. We require the 2-dimensional Fourier transform $F(\rho; \phi)$ of $f(r; 0) = f(x, y)$. After lengthy, but nevertheless straightforward, manipulations it transpires that

$$F_s(\rho; \phi) = (\pi^{1/2}\theta/2^{3/2}) \times \exp(-2\pi^2R^2\theta^2\rho^2) \int_{-\infty}^{\infty} \exp(-t^2/2\theta^2) \times F([1 + \tau^2]^{1/2}\rho; \phi - \tan^{-1}(\tau))\, dt \quad (10)$$

where $\tau = t - j\bar{\mu}/\rho - j2\pi R\theta^2\rho$. In situations where it is allowable to neglect the detector's beamwidth (i.e. $\theta \approx 0$) we see that the Gaussian function in the integrand has the character of a delta function, implying that eqn. 10 reduces to

$$F_s(\rho; \phi) = F((\rho^2 - \bar{\mu}^2)^{1/2}; \phi + j \tanh^{-1}(\bar{\mu}/\rho)) \quad (11)$$

which emphasises Clough and Barrett's [8] elegant insight that useful information concerning the conventional spectrum $F(\rho; \phi)$ is given only by the part of the SPECT spectrum $F_s(\rho; \phi)$ lying outside the circle of radius $\bar{\mu}$ centred on the origin of Fourier space. It is unclear how to make practical computational use in general of values of $F_s(\rho; \phi)$ for $\rho < \bar{\mu}$, because these values define $F(\rho; \phi)$ for imaginary values of its radial co-ordinate. The fact that $F_s(\rho; \phi)$ defines $F(\rho; \phi)$ explicitly only for complex values of ϕ does not mean that $F(\rho; \phi)$ cannot be readily evaluated for real values of ϕ. Inspection of eqn. 11 reveals that, when $F_s(\rho; \phi)$ and $F(\rho; \phi)$ are both expressed as trigonometrical Fourier series in ϕ, their corresponding Fourier coefficients are directly proportional to each other, with the quantity $\tanh^{-1}(\bar{\mu}/\rho)$ appearing in the proportionality factors.

We are interested in situations in which the factor $\exp(-t^2/2\theta^2)$, in eqn. 10, does not have the character of a delta function. The form of eqn. 10 is then decidedly discouraging, and we have found it more profitable to return to eqn. 9.

Various practical difficulties cause the resolution achievable with SPECT systems to be quite limited [14, 19]. It is, consequently, adequate to represent $f(x, y)$ as an array of 'Gaussian blobs'. The spacing of the blobs and the effective radius, σ say, of each blob characterise the resolution which one hopes a particular SPECT system will achieve.

Because $f(x, y)$ is linearly related to either $p_s(\xi; \phi)$ or $F_s(\rho; \phi)$, the performance of a SPECT system can be

assessed by examining a single blob, centred at the point $(b; \beta)$ say, i.e.

$$f(x, y) = \exp\{-[(x - b\cos(\beta))^2 + (y - b\sin(\beta))^2]/2\sigma^2\} \quad (12)$$

When eqn. 12 is substituted into eqn. 9, both integrations can be performed immediately (but care must be taken when manipulating the integrands never to allow η to become implicitly less than $(-R)$, as is clear from inspection of Fig. 1), giving

$$F_s(\rho; \phi) = H(\rho; \phi, \bar{\mu}, \sigma, R, b; \beta; 0)F(\rho; \phi, b; \beta) \quad (13)$$

where $F(\rho; \phi, b; \beta)$ is the spectrum of the single blob, i.e.

$$F(\rho; \phi, b; \beta) = 2\pi\sigma^2 \exp(-2\pi[\pi\sigma^2\rho^2 - jb\rho\cos(\phi - \beta)]) \quad (14)$$

and $H(\rho; \phi, \bar{\mu}, \sigma, R, b; \beta; 0)$ represents the spatial frequency distortion due to both (constant) attenuation of the radiations and (variable) finite resolution of the sources of these radiations, i.e.

$$H(\rho; \phi, \bar{\mu}, \sigma, R, b; \beta; 0) = (\theta^2/2\omega)\exp(-\Omega) \quad (15)$$

where

$$\omega = (1 + 4\pi^2\sigma^2\theta^2\rho^2)^{1/2} \quad (16)$$

and

$$\Omega = 2\pi\{\pi[1 - 4\pi\bar{\mu}\sigma^2/R]R^2\theta^2\rho^2 - \pi\bar{\mu}^2\sigma^2 - \bar{\mu}b\sin(\phi - \beta) - 2\pi bR\theta^2\rho^2\sin(\phi - \beta) + \pi b^2\theta^2\rho^2\sin^2(\phi - \beta)\}/\omega^2 \quad (17)$$

Eqns. 12 to 17 suggest a simple image-processing scheme for SPECT, as we explain in the following two Sections. Before embarking on this, we point out that $F_s(\rho; \phi)$ is not, as it stands, a suitable basis for image reconstruction because it is not conjugate symmetric, i.e. $F_s(\rho; \phi + \pi) \neq F_s^*(\rho; \phi)$, where the asterisk denotes complex conjugation, implying that the image is not real. It is, therefore, appropriate to introduce what we call the oppositely averaged SPECT spectrum $\bar{F}_s(\rho; \phi)$ defined by

$$\bar{F}_s(\rho; \phi) = [F_s(\rho; \phi) + F_s(-\rho; \phi + \pi)]/2 \quad (18)$$

which is conjugate symmetric. Note that using $\bar{F}_s(\rho; \phi)$ as a basis for image reconstruction is equivalent to averaging diametrically opposed projections, which is a standard SPECT procedure [5].

5 Resolution considerations

To generate useful SPECT imagery, the resolution must be much less than the average diameter of the cross-section. So, we must insist that $\sigma \ll R$. The term in Ω, as defined by eqn. 17, which most limits the spatial frequency content of the SPECT spectrum is $2\pi^2R^2\theta^2\rho^2$. It follows that ω, as defined by eqn. 16, differs little from unity throughout the range of ρ for which $\exp(-2\pi^2R^2\theta^2\rho^2)$ has appreciable value. Consequently, ω can be replaced by unity in eqn. 17 without introducing any significant error.

If $\bar{\mu}$ is large enough that the attenuation is appreciable over a distance equal to the effective radius of the blob, then either this radius is set too large or the magnitude of the attenuation is so severe that the SPECT system is unlikely to be of much practical use. Consequently, the factor $2\pi\bar{\mu}\sigma$ must be appreciably less than unity, say no greater than 1/3. Therefore, $(4\pi^2\bar{\mu}\sigma^2/R)$ can be no greater than $2\sigma/3R$, which is itself considerably less than unity (see the second sentence of this Section). Similarly, $2\pi\bar{\mu}^2\sigma^2$ cannot exceed 1/18. We then see that eqns. 15 to 17 reduce effectively to

$$H(\rho; \phi, \bar{\mu}, \sigma, R, b; \beta; 0) = (\theta^2/2)\exp(-2\pi^2R^2\theta^2\rho^2 - \Lambda) \quad (19)$$

where

$$\Lambda = 2\pi^2\{\sin^2(\phi - \beta) - [2R/b + \bar{\mu}/\pi b\theta^2\rho^2] \times \sin(\phi - \beta)\}b^2\theta^2\rho^2 \quad (20)$$

from which, it appears that attenuation only contributes appreciably to image degradation at low spatial frequencies. We must not, of course, forget the normalisation involved in the definition of $h(P, Q)$ in eqn. 7. This normalisation glosses over the very significant adverse effect of the attenuation on the detected signal-to-noise ratio. Because the maximum value of b can hardly be greater than $2R/3$, say, in practice, and is likely to be closer to $R/2$, we see from eqns. 13, 14, 18, 19 and 20 that any blob's oppositely averaged SPECT spectrum is dominated by the factor

$$M(\rho) = (\pi\sigma^2\theta^2)\exp[-2\pi^2(\sigma^2 + R^2\theta^2)\rho^2] \quad (21)$$

which determines the resolution which one should attempt to achieve in any particular instance. The point is that $|\bar{F}_s(\rho; \phi)|$ is only appreciable in that part of Fourier space wherein $M(\rho)$ exceeds a threshold set by the prevailing noise level.

The total flux of radiation received from a blob is proportional to $(\sigma\theta)^2$, with the constant of proportionality depending on the various system parameters such as the species and densities of the sources of radiation, the time devoted to measuring each SPECT projection and the nature of the detectors. Because $M(0)$ is directly proportional to this flux, as is $1/R^2$, the ratio of $M(\rho)$ to $\pi k^2/R^2$ (where the constant k depends on the system parameters) characterises the signal-to-noise ratio for the component of the spectrum at the spatial frequency ρ. As the effective radius of each blob is σ, we require $M(\rho)$ to exceed the lowest spectral level that can be determined to a useful accuracy for all ρ up to $1/\sigma$. Taking $\pi k^2/R^2$ to be this level, we see that

$$M(1/\sigma) = \pi k^2/R^2 \quad (22)$$

Combining eqns. 21 and 22 gives

$$\alpha^2 = \pi^2/(q + \log_e \alpha) \quad (23)$$

where $\alpha = (\sigma/R\theta)$ and $q = \log_e(R^2\theta^2/k) - \pi^2$.

In any particular instance, the system parameters are known and so eqn. 23 can be solved for α. This determines the system resolution because it fixes the choice of σ.

One can attempt to remove some, small, part of the blurring inherent in $H(\rho; \phi, \bar{\mu}, \sigma, R, b; \beta; 0)$ by invoking an image restoration strategy of the kind referenced in Section 1. As such strategies are covered adequately in the quoted References, we content ourselves with propounding a simple enhancement scheme which is suggested by the analytical approach introduced in this paper.

On recalling that the resolution is essentially determined by $M(\rho)$, we see, from eqns. 13 to 20, that the only factor in $\bar{F}(\rho; \phi)$ that depends, effectively, on $\bar{\mu}$ is

$$J(\phi, \bar{\mu}, b; \beta) = \cosh[2\pi b\bar{\mu}\sin(\phi - \beta)] \quad (24)$$

the practical computational significance of which is conveniently assessed by examining the multi-blob image shown in Fig. 2. This is an idealised $f(x, y)$ consisting of a discrete array of identical blobs. Fig. 3 shows the reconstruction, directly from $\bar{F}_s(\rho; \phi)$. The attenuation coefficient is typical of that occurring in practice. Although the

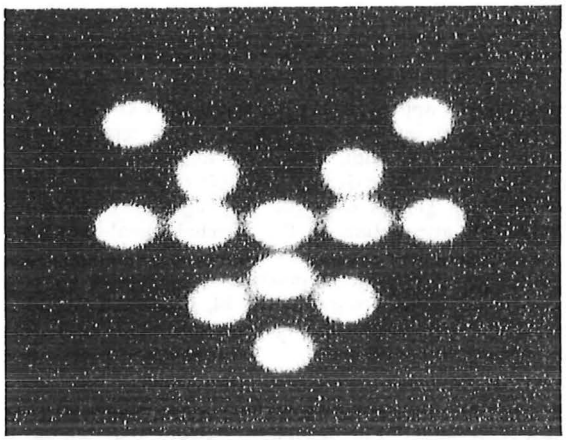

Fig. 2 *Multi-blob image, consisting of 13 identical blobs ($\sigma = 0.08$)*
The co-ordinates of the centres of the blobs are (0.6; π), (0.3; π), (0.0; 0), (0.3; 0), (0.6; 0.0), (0.4; $\pi/4$), (0.8; $\pi/4$), (0.4; $3\pi/4$), (0.8; $3\pi/4$), (0.5; $4\pi/3$), (0.5; $5\pi/3$), (0.3; $3\pi/2$), (0.7; $3\pi/2$)

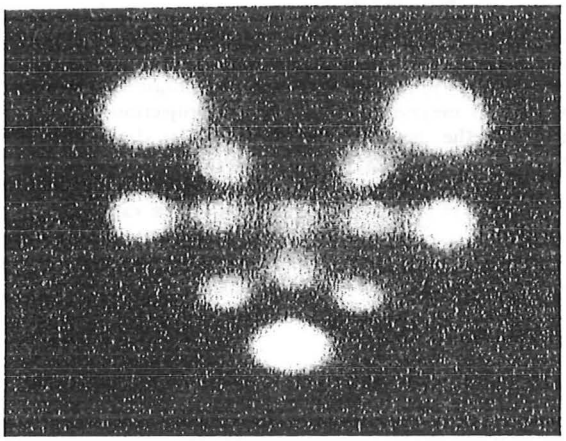

Fig. 3 *Magnitude of uncompensated reconstruction of multi-blob image, assuming $R = 1$, $\theta = 0.05$ and $\bar{\mu} = 0.5$*

shapes of the reconstructed blobs are somewhat distorted, their most serious defect is the large variation in their amplitudes. The enhancement strategy described in the following Section compensates for much of this.

6 Simple image enhancement strategy

The variation in the amplitudes of the blobs apparent in Fig. 3 is mainly due to the factor $J(\phi, \bar{\mu}, b; \beta)$ defined by eqn. 24. The enhancement strategy introduced here is based on the following interpretation of this factor. We first note, from the definition of the Fourier transform [15], that

$$f(0, 0) = \int_0^{\infty} \int_0^{2\pi} F(\rho; \phi)\rho \, d\phi \, d\rho \tag{25}$$

which suggests that the aforementioned variation in blob amplitudes may be given to a useful degree of approximation by the integral of $J(\phi, \bar{\mu}, b; \beta)$ over the ρ, ϕ-plane. As $J(\phi, \bar{\mu}, b; \beta)$ is independent of ρ, we need only consider the factor

$$A(b, \bar{\mu}) = \int_0^{2\pi} J(\phi, \bar{\mu}, b; \beta) \, d\phi \tag{26}$$

By expanding $J(\phi, \bar{\mu}, b; \beta)$ as a series of modified Bessel functions [20], we obtain

$$A(b, \bar{\mu}) = A(\bar{\mu}b) = I_0(2\pi\bar{\mu}b) \tag{27}$$

the form of which is displayed in Fig. 4, for the interval $0 \leqslant \bar{\mu}b \leqslant 0.5$.

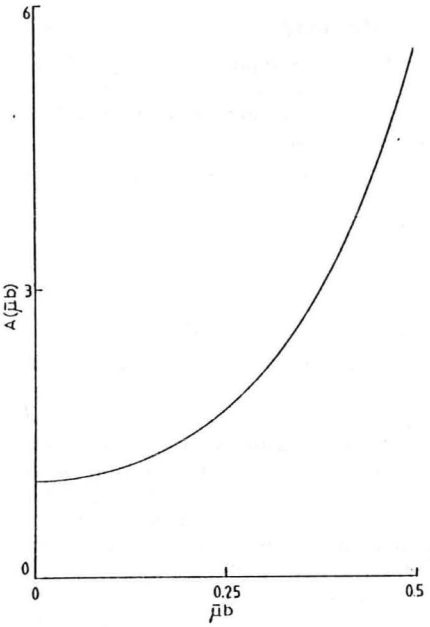

Fig. 4 *Radial amplitude variation factor $A(\bar{\mu}b)$*

The enhancement procedure is to take the image obtained directly from $\bar{F}_s(\rho; \phi)$ and, at each point, divide the value of the image by $A(\bar{\mu}r)$. Fig. 5 reveals the result of enhancing the image shown in Fig. 3. While this

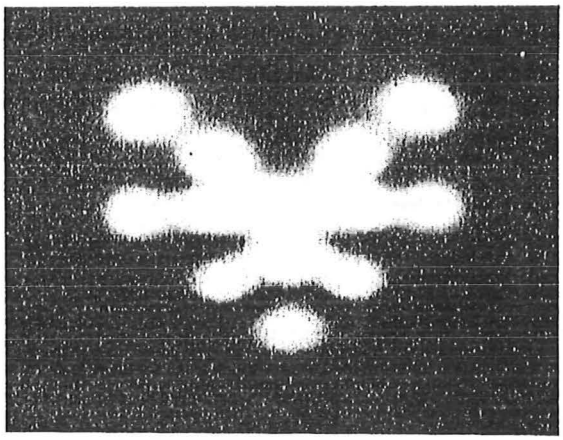

Fig. 5 *Magnitude of enhanced reconstruction of multi-blob image*
Same parameters as for Fig. 3

enhancement procedure is certainly approximate, we find it to be effective over a wide range of values of $\bar{\mu}$. Blobs at different radii are reconstructed to somewhat different resolutions, but the variations are likely to be acceptable in practice. The amplitude variations with radius are reduced markedly. We demonstrate this quantitatively with the aid of the 5-blob image, shown in Fig. 6, which

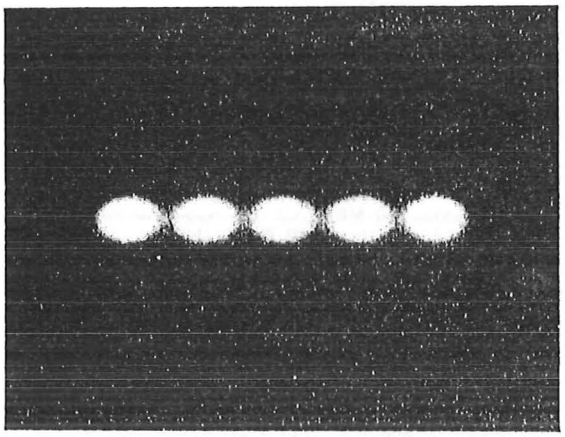

Fig. 6 *5-blob image, consisting of identical blobs ($\sigma = 0.08$) all centred on x-axis at points defined by $x = 0, \pm 0.3, \pm 0.6$*

consists of a linear discrete array of blobs each identical to each of those shown in Fig. 2. Fig. 7 shows the reconstruction of the 5-blob image, directly from $\bar{F}_s(\rho; \phi)$, without any kind of restoration being attempted. The

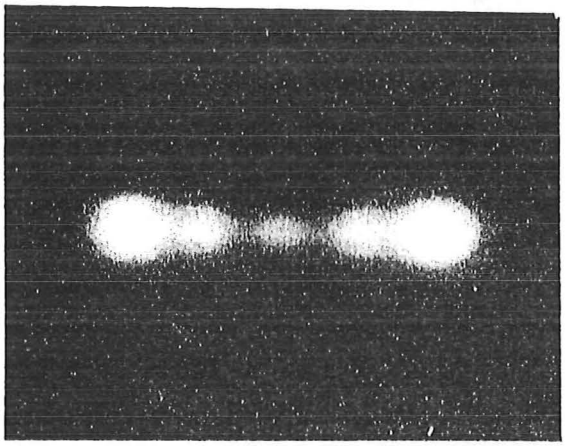

Fig. 7 *Magnitude of uncompensated reconstruction of 5-blob image, assuming $R = 1$, $0 = 0.05$ and $\bar{\mu} = 0.5$*

attenuation coefficient is the same as that invoked for Fig. 3. When the image shown in Fig. 7 is enhanced, it is quite similar to Fig. 6, although there is a significant loss of resolution. Fig. 8 shows line profiles along the x-axis through the original image and also through the uncompensated and enhanced images. The main differences between the profiles through the original and enhanced images are that the latter profile exhibits slightly broader peaks of somewhat unequal amplitudes and its nulls are less than half as deep.

Our suggested image-processing strategy is thus to record a set of measured projections within the angular range $0 \leqslant \phi < 2\pi$, normalise each projection by $h(P, Q)$ as defined by eqn. 7, thereby generating a set of what we call (in Section 3) SPECT projections, average the SPECT projections at angles differing by π, thereby generating a set of 'averaged' projections within the angular range $0 \leqslant \phi < \pi$, filter each of these averaged projections such that their spatial frequency content corresponds to the resolution estimated by the procedure characterised by eqns. 22 and 23, apply a conventional CT image reconstruction algorithm (refer to final paragraph of Section 3), and, finally, divide each pixel value in the reconstructed image (whose radial co-ordinate is r) by $A(\bar{\mu}r)$, where $2\pi\bar{\mu}$ is the available estimate of the average attenuation coefficient. It is the combination of the filtering and the division by $A(\bar{\mu}r)$ which represents the outcome of the theoretical development introduced in this paper.

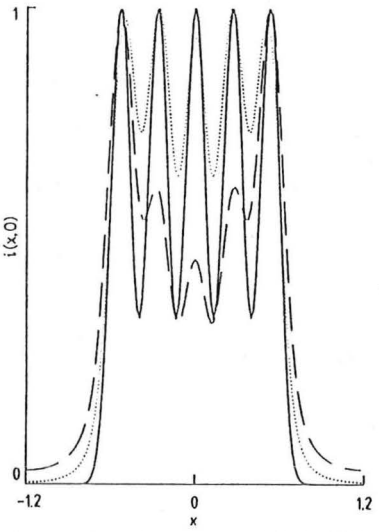

Fig. 8 *Line profiles along x-axis for the 5-blob image*

——— Fig. 6
- - - - Fig. 7
..... magnitude of enhanced image

7 Effects of variable resolution

When, as is always true in practice, the attenuation is variable, the SPECT spectrum of a blob is given (approximately, but nevertheless to reasonable accuracy) by $F_s(\rho; \phi)$, as developed in Sections 4 and 5, multiplied by a factor $Z(\rho; \phi, b; \beta)$ which of course depends, in general, on where the blob is located in the cross-section. As this factor represents only the variation of the attenuation about its average value, it can usually be expected to merely affect the resolution to which the blob is reconstructed. We remark that this factor has an identical form to that which arises when the amplitude of the image $f(x, y)$ varies from projection to projection. We have previously studied such 'temporal variations' (because projections are usually measured sequentially, the aforesaid variations are usefully thought of as occurring in time) and have shown that they have surprisingly little effect on image detail [21, 22]. Boundaries between regions of different 'density' tend to be clearly recognisable, even when the variations are large [21].

The main purpose of a SPECT system is to identify locations of appreciable concentrations of sources of radiation. One seldom attempts to recover fine detail (which is just as well, considering the relatively poor

resolutions and signal-to-noise ratios of SPECT systems). We therefore feel confident in asserting, on the basis of our previous experience [21, 22], that only rarely will be effects of variable resolution be actually noticeable in images reconstructed from SPECT projections.

We are not, of course, inferring that attenuation has no important effects. It seriously degrades the signal-to-noise ratio, as we have already emphasised. Compensation for the average value of the attenuation coefficient is mandatory, unless one is prepared to put up with image degradation of the kind illustrated in Figs. 3 and 7. We have, however, shown in Section 6 how to ameliorate the effects of average attenuation. What we are arguing in this Section is that variable attenuation is usually comparatively unimportant.

8 Conclusions

Our combined analysis of variable resolution and constant attenuation in SPECT has persuaded us that:

(*a*) variable resolution is conveniently handled by representing the image as a distribution of blobs whose effective diameters are chosen according to the simple prescription introduced in Section 5

(*b*) the only really serious image degradation due to constant attenuation (apart from its adverse effect on the signal-to-noise ratio) is its tendency to cause different parts of the image to be reconstructed with different amplitudes (this tendency can be markedly inhibited by the simple enhancement procedure introduced in Section 6)

(*c*) it is probably a waste of time to attempt to solve the variable attenuation problem.

We are planning experiments to evaluate (in a clinically meaningful manner) the usefulness of the image-processing strategy suggested in Section 6 and to compare it with more sophisticated processing, such as that recently reported by Faber *et al.* [23] and Webb [24].

9 Acknowledgments

D.G.H. Tan acknowledges the award of a Postgraduate Research Scholarship from the New Zealand University Grants Committee, while J.X. Qu thanks the University of Science and Technology of China, for providing financial support, and the Ministry of Foreign Affairs of New Zealand, for making the necessary arrangements for his visit. K.L. Garden and R.H.T. Bates thank Technicare Corporation of Cleveland, Ohio, for financial assistance for our research. We are especially grateful for technical correspondence with George Kambic and Steve Gottschalk. R.H.T. Bates thanks Anne Clough for helpful discussions during an early stage of the work reported here.

10 References

1 PETERS, T.M.: 'Image reconstruction from projections'. Ph.D. thesis, Engineering Library, University of Canterbury, Christchurch, New Zealand, 1973

2 GULLBERG, G.T., and BUDINGER, T.F.: 'The use of filtering methods to compensate for constant attenuation in single-photon emission computed tomography', *IEEE Trans.*, 1981, **BME-28**, pp. 142–157

3 BATES, R.H.T., GARDEN, K.L., and PETERS, T.M.: 'Overview of computerized tomography with emphasis on future developments', *Proc. IEEE*, 1983, **71**, pp. 356–372

4 GULLBERG, G.T.: 'The attenuated Radon transform: theory and application in medicine and biology'. Ph.D. thesis (Tech. Rep. LBL-7486), Lawrence Berkeley Lab., University of California, 1979

5 GREER, K.L., JASZCZAK, R.J., and COLEMAN, R.E.: 'An overview of a camera-based SPECT system', *Med. Phys.*, 1982, **9**, pp. 455–463

6 GULLBERG, G.T.: 'The attenuated Radon transform: application to single photon emission computed tomography in the presence of a variable attenuating medium'. Tech. Rep. LBL-10276, Lawrence Berkeley Lab., University of California, 1980

7 TRETIAK, O.J., and METZ, C.: 'The exponential Radon transform', *SIAM J. Appl. Math.*, 1980, **39**, pp. 321–335

8 CLOUGH, A.V., and BARRETT, H.H.: 'Attenuated Radon and Abel transforms', *J. Opt. Soc. Am.*, 1983, **73**, pp. 1590–1595

9 GARDEN, K.L.: 'An overview of computed tomography'. Ph.D. thesis, Engineering Library, University of Canterbury, Christchurch, New Zealand, 1984

10 HSIEH, R.C., and WEE, W.G.: 'On methods of three-dimensional reconstruction from a set of radioisotope scintigrams', *IEEE Trans.*, 1976, **SMC-6**, pp. 854–862

11 ANSARI, A., and WEE, W.G.: 'Reconstruction from projections in the presence of distortion'. *IEEE Conf. Proc. Decision & Control* (New Orleans), 1977, **1**, pp. 361–366

12 FLOYD, C.E., JASZCZAK, R.J., GREER, K.L., and COLEMAN, R.E.: 'Deconvolution of Compton scatter in SPECT', *J. Nucl. Med.*, 1985, **26**, pp. 403–408.

13 MADSEN, M.T., and PARK, C.H.: 'Enhancement of SPECT images by Fourier filtering the projection image set', *J. Nucl. Med.*, 1985, **26**, pp. 395–402

14 LIM, J.S.: 'Performance analysis of three camera configurations for single photon emission computed tomography', *IEEE Trans.*, 1980, **NS-2**, pp. 559–568

15 LEWITT, R.M.: 'Reconstruction algorithm: transform methods', *Proc. IEEE*, 1983, **71**, pp. 390–408

16 GOTTSCHALK, S.C., SALEM, D., LIM, C.B., and WAKE, R.H.: 'SPECT resolution and uniformity improvements by noncircular orbit', *J. Nucl. Med.*, 1983, **24**, pp. 822–828

17 CHANG, L.T.: 'Attenuation correction and incomplete projection in single photon emission computed tomography', *IEEE Trans.*, 1979, **NS-26**, pp. 2780–2789

18 DEUTSCH, M.: 'New inverses of the attenuated Abel integral equation', *J. Phys. A*, 1984, **17**, pp. L939–L944

19 JASZCZAK, R.J., COLEMAN, R.E., and WHITEHEAD, F.R.: 'Physical factors affecting quantitative measurements using camera-based single photon emission computed tomography (SPECT)', *IEEE Trans.*, 1981, **NS-28**, pp. 69–80

20 ABRAMOWITZ, M., and STEGUN, I.A. (Eds.): 'Handbook of mathematical functions' (Dover, New York, 1965), p. 376

21 GARDEN, K.L., BATES, R.H.T., WON, M.C., and CHIKWANDA, H.: 'Computerized tomographic imaging is insensitive to density variation during scanning', *Image & Vision Comput.*, 1984, **2**, pp. 76–84

22 GARDEN, K.L., and BATES, R.H.T.: 'Image reconstruction from projections VII: interactive reconstruction of piecewise constant images from few projections', *Optik*, 1984, **68**, pp. 161–173

23 FABER, T.L., LEWIS, M.H., CORBETT, J.R., and STOKELY, E.M.: 'Attenuation correction for SPECT: An evaluation of hybrid approaches', *IEEE Trans.*, 1984, **MI-3**, pp. 101–107

24 WEBB, S.: 'Comparison of data-processing techniques for the improvement of contrast in SPECT liver tomograms', *Phys. Med. & Biol.*, 1985, **30**, pp. 1077–1086

6-B NEWTON-KANTOROVICH METHOD APPLIED TO INVERSE SCATTERING FOR AN EXTERIOR HELMHOLTZ PROBLEM

This appendix is a preprint of a paper (authors: D. G. H. Tan and D. J. N. Wall) submitted (January 1988) to *Inverse Problems.*

Abstract

The problem of determining the shape of an impenetrable obstacle from a set of measurements of its far-field scattering amplitude is solved by formulating the problem as a non-linear operator equation which is then solved by an iterative method. The use of the null-field method to solve the direct problem at each iteration leads to efficient evaluation of the Fréchet derivative of the non-linear operator. Results of numerical trials with simulated data are presented.

1 INTRODUCTION

Inverse problems arise in a large number of important technological and scientific applications [1,2,3] and are often conveniently formulated in terms of an associated direct problem which has received much more attention [4]. Such is the case when wave motion involving diffraction is concerned. The direct problem, relevant to the inverse problem considered in this paper, consists of knowledge of a scattering obstacle embedded in an otherwise homogeneous medium, and specification of an incident field describing radiation impinging on the scatterer. Equations defining the interaction of the incident field with the obstacle must then be solved to deduce the resultant scattered field. The associated inverse problem is to deduce the form of the scatterer based on knowledge of the scattered fields caused by a number of incident fields. Once it is realised that the direct problem defines a mapping $\mathcal{M}$ from the space C of scatterers to the space E of measurements,

$$\mathcal{M} : C \rightarrow E, \tag{1}$$

the inverse problem is readily posed as finding the inverse mapping

$$\mathcal{M}^{-1} : E \rightarrow C. \tag{2}$$

At once, one is tempted to ask whether the inverse map exists and is unique, and whether the solution is stable [4]. It is in general difficult to answer these questions because $\mathcal{M}$ is normally non-linear and often cannot be written in closed form, but see Colton [5] for some discussion of these questions.

The inverse problem becomes greatly simplified if $\mathcal{M}$ can be meaningfully approximated by a linear map. Techniques such as computed tomography [6], which has revolutionised clinical medical diagnostics, and diffraction tomography [7] are popular inversion schemes because their linear models of the direct problem are readily inverted numerically.

When a linear approximation to $\mathcal{M}$ is deemed invalid, the only recourse is to attempt indirect iterative inversion of the non-linear expression (1). One technique that has proved particularly useful is the Newton-Kantorovich method [8,9,10]. In this paper, we apply the Newton-Kantorovich method to an inverse problem whose associated direct problem is an exterior Helmholtz equation. The obstacle is such that the Dirichlet boundary condition applies on its surface, as is appropriate for the tangential electric field on a perfectly conducting obstacle, or for a sound soft acoustic obstacle. This problem has been considered before [9,10,11] and is known to be ill-posed [5]. What is original about the results reported here is that our formulation of the direct problem employs the null-field method [12] in a manner that enables simple, efficient, and stable numerical evaluation of the Fréchet derivative as required by the Newton-Kantorovich method.

The remainder of this paper is organised as follows. In Section 2 we introduce our notation and describe a numerical method which serves to define the direct mapping $\mathcal{M}$. Section 3 describes the basic principles of the Newton-Kantorovich method and the Fréchet derivative. Equations defining the Fréchet derivative for our direct mapping are presented in Section 4. Consideration of the numerical implementation of the Newton-Kantorovich method is shown in Section 5 to lead to a stable and efficient numerical scheme for evaluation of the Fréchet derivative. Numerical results verifying our analysis are also presented in Section 5 and in Section 6 we assess the significance of our results.

2 DIRECT PROBLEM

Our technique developed in this paper applies in two or three spatial dimensions but, in order to convey the essential principles without becoming obscured in notational nomenclature, the exposition considers a problem depending on only two spatial dimensions. An arbitrary point in the plane Υ is denoted in Cartesian coordinates by (x, y) and in polar coordinates by (r, θ). The concepts discussed here are easily generalised to three dimensions. A smooth, simply connected, closed curve Γ partitions Υ into an interior region Υ_-, containing the origin O, and an exterior region Υ_+ which is assumed to be homogeneous with wavespeed c. It is assumed that Γ is star-like, i.e. an arbitrary point on Γ can be represented in polar coordinates as (ρ, ϕ) where ρ is a real single-valued function of ϕ,

$$\rho = \rho(\phi). \tag{3}$$

Time-harmonic waves with the suppressed temporal dependence $\exp(i\omega t)$ propagate in the exterior region and the presence of an impenetrable obstacle within Υ_- causes a scattered field ψ_s that is described by

$$\begin{aligned}
(\nabla^2 + k^2)\psi_s(\vec{x}, k) &= 0, \qquad \vec{x} \in \Upsilon_+, && (4)\\
\lim_{r\to\infty} r^{1/2}(\partial\psi_s/\partial r + ik\psi_s) &= 0, && (5)\\
\psi &= \psi_i + \psi_s, \quad \vec{x} \in \Upsilon_+ \cup \Gamma, && (6)\\
\psi &= 0, \qquad \vec{x} \in \Gamma, && (7)
\end{aligned}$$

where $k = \omega/c = 2\pi/\lambda$ is the wavenumber and λ is the wavelength of the radiation. Condition (5) represents the Sommerfeld radiation condition [13]. The Dirichlet boundary condition (7) is known as the "sound soft" boundary condition in acoustics, where ψ represents the velocity potential, and is also appropriate when ψ represents the axially directed component of electric field on a perfect conductor. The incident field ψ_i also satisfies (4) throughout Υ_+, except for the region containing the source of the radiation. The source is assumed entirely under the control of the experimenter and is of no further concern.

It suffices to assume that ψ_i can be chosen to be a plane wave, satisfying (4) exactly throughout the whole of Υ, with the multipole expansion [14, page 1371]

$$\begin{aligned}
\psi_i &= \exp(-i\vec{k}\cdot\vec{x})\\
&= \sum_{m=-\infty}^{\infty} a_m J_m(kr)\exp(im\theta), && (8)
\end{aligned}$$

where J_m denotes the Bessel function of the first kind of order m, $\vec{k}$ is a vector in the plane satisfying $|\vec{k}| = k$, and $a_m = (-i)^m \exp(im\Theta)$, Θ being the angle formed by $\vec{k}$ and the positive x-axis.

A technique to compute ψ_s, for a given Γ and ψ_i, popular among numerical practitioners and with considerable theoretical advantages, is the null-field method [12] which is based on the coupled integral equations

$$\oint_\Gamma g(|\vec{x} - \vec{x}'|, k)\frac{\partial u}{\partial n}(\vec{x}', k)\, d\Gamma(\vec{x}') = \begin{cases} \psi_i(\vec{x}, k), & \vec{x} \in \Upsilon_-,\\ -\psi_s(\vec{x}, k), & \vec{x} \in \Upsilon_+. \end{cases} \tag{9}$$

Here, g denotes the free-space Green's function $-iH_0^{(2)}(k|\vec{x} - \vec{x}'|)/4$, where $H_m^{(2)}$ is the Hankel function of the second kind of order m, and $\partial/\partial n$ denotes the directional derivative in the direction of the outward normal to Γ. On using (3) to parametrise Γ,

and expanding g with the aid of an addition theorem [15, §11.3], the Petrov-Galerkin method with "test" functions

$$\varphi(\vec{x}, k) = 4iC_m \exp(im\theta), \quad m = -\infty, \ldots, \infty, \tag{10}$$

applied to (9) with $\vec{x} \in \Upsilon_-$ results in an infinite set of moment equations. The set $\{C_m\}$ are constants to be chosen later. The null-field equations so obtained are

$$C_m \int_0^{2\pi} f(\phi) H_m^{(2)}(k\rho(\phi)) \exp(-im\phi) d\phi = U_m, \quad m = -\infty, \ldots, \infty, \tag{11}$$

where $U_m = 4iC_m a_m$ and the surface source density f depends on Γ through

$$f(\phi) = \sqrt{[\rho(\phi)]^2 + (d\rho/d\phi)^2}\, \frac{\partial u}{\partial n}(\rho(\phi), \phi, k). \tag{12}$$

Since f has period 2π, it is appropriately expanded as

$$f(\phi) = \sum_{m=-\infty}^{\infty} f_m \exp(im\phi), \tag{13}$$

so that the null-field equations reduce to solving the infinite system of linear equations

$$\sum_{n=-\infty}^{\infty} H_{mn} f_n = U_m, \quad m = -\infty, \ldots, \infty, \tag{14}$$

where, for $m, n = -\infty, \ldots, \infty$,

$$H_{mn} = C_m \int_0^{2\pi} H_m^{(2)}(k\rho(\phi)) \exp(i(n-m)\phi)\, d\phi. \tag{15}$$

In the far-field, $\vec{x} \in \Upsilon_+$ is sufficiently far removed from Γ to permit the asymptotic form of the Hankel functions to be invoked in the second equation of (9), so that the scattered field has the asymptotic form

$$\psi_s \sim (1/8\pi kr)^{-1/2} \exp(-i(kr - 3\pi/4)) \sum_{m=-\infty}^{\infty} b_m i^m \exp(im\theta), \tag{16}$$

where, from (9) with $\vec{x} \in \Upsilon_+$, the b_m are given by

$$b_m = \sum_{n=-\infty}^{\infty} J_{mn} f_n, \tag{17}$$

and

$$J_{mn} = \int_0^{2\pi} J_m(k\rho(\phi)) \exp(i(n-m)\phi)\, d\phi. \tag{18}$$

Note that since $J_m = \mathrm{Re}(H_m^{(2)})$, where Re denotes the real part, the integrals in (18) will have been computed in the evaluation of (15).

When solving the direct problem for a particular $\rho(\phi)$, the conditioning of the linear system (14) is improved by suitable normalisation of the Hankel functions appearing in (15) [16]. Equivalently, this may be accomplished by choosing the constants C_m to be $1/H_m^{(2)}(k\rho(0))$. Thus, in the event that Γ is a circle of radius ρ_0, H_{mn} reduces to the Kronecker delta, and [10]

$$f_m = (-i)^m / H_m^{(2)}(k\rho_0), \tag{19}$$

$$b_m = (-i)^m J_m(k\rho_0) / H_m^{(2)}(k\rho_0), \tag{20}$$

as can be deduced from separation of variables [17, §2.2.1].

When (7) is replaced by the Neumann boundary condition,

$$\partial\psi/\partial n = 0, \quad \vec{x} \in \Gamma, \tag{21}$$

it is found that (12), (15), and (18) are replaced by

$$f(\phi) = u(\rho(\phi), \phi, k), \tag{22}$$

$$H_{mn} = -C_m \int_0^{2\pi} \left[k\rho(\phi) H_m^{(2)\prime}(k\rho(\phi)) + im\rho'(\phi) H_m^{(2)}(k\rho(\phi))\right] \exp(i(n-m)\phi)\, d\phi, \tag{23}$$

$$J_{mn} = -\int_0^{2\pi} [k\rho(\phi) J_m'(k\rho(\phi)) + im\rho'(\phi) J_m(k\rho(\phi))] \exp(i(n-m)\phi)\, d\phi. \tag{24}$$

In (24) and (23), the prime denotes differentiation with respect to the argument of the appropriate functions.

3 GENERAL NEWTON-KANTOROVICH ALGORITHM

The Newton-Kantorovich algorithm [18] is an iterative scheme to solve non-linear operator equations whose most familiar manifestation is, perhaps, Newton's solution of non-linear algebraic equations. Consider a general non-linear operator equation

$$\mathcal{M}(x) = \mathcal{D}, \tag{25}$$

with x belonging to a normed linear space. At each iteration, $\mathcal{M}$ is linearised about the current estimate $x^{(l)}$ of a solution,

$$\mathcal{M}(x^{(l)} + \delta x^{(l)}) \approx \mathcal{M}(x^{(l)}) + \mathcal{M}'(x^{(l)})\delta x^{(l)}, \tag{26}$$

where $\mathcal{M}'$ denotes the Fréchet derivative of $\mathcal{M}$. To compute an improved estimate, (26) is equated to the right hand side of (25), and the resulting equation

$$\mathcal{M}'(x^{(l)})\delta x^{(l)} = \mathcal{D} - \mathcal{M}(x^{(l)}) \tag{27}$$

is solved (in a least squares sense if over-determined) for the update $\delta x^{(l)}$. The new estimate of a solution is thus

$$x^{(l+1)} = x^{(l)} + \lambda^{(l)}\delta x^{(l)}, \tag{28}$$

where, for the conventional Newton-Kantorovich algorithm, $\lambda^{(l)}$ is always unity. Improved convergence to a solution may be obtained by performing a line search to determine an optimal $\lambda^{(l)}$ at each iteration. Techniques such as the Levenberg-Marquardt algorithm [19] may also improve the algorithm. An important factor in the success of the basic method is the faithfulness of the initial approximation $x^{(0)}$.

The Newton-Kantorovich formulation provides a general method which can be employed in many applications. The data for a given inverse problem represent $\mathcal{D}$. In this paper, the data constitute measurements of the scattered field. The form of $\mathcal{M}'$ depends, of course, on $\mathcal{M}$, and the success of the method depends on reliable calculation, at each $x^{(l)}$, of both of these quantities. For the direct problem considered here, the mapping of a particular obstacle to its scattered field is computed using the null-field equations of the previous section. It remains to describe, in the next section, a method for computing the Fréchet derivative of the direct mapping.

4 COMPUTATION OF FRÉCHET DERIVATIVE

The operator $\mathcal{M}$ considered here is defined as a set of normalised measurements of the scattered field due to a single incident wave, made in the far-field at a number of different angles, denoted θ_{meas}. This mapping is conveniently expressed by the functional

$$\mathcal{M}(\rho) = \sum_{m=-\infty}^{\infty} i^m b_m \exp(im\theta_{\text{meas}}), \tag{29}$$

where the functional dependence on $\rho(\phi)$ is implicit in $\{b_m\}$ through (14), (15), (17), and (18). It is appropriate to introduce a representation for Γ in terms of some basis coefficients $\{\rho_p\}$ and basis functions $\{\varrho_p(\phi)\}$,

$$\rho(\phi) = \sum_p \rho_p \varrho_p(\phi), \tag{30}$$

where, for the moment, the number of basis functions and their functional form is completely arbitrary. The Fréchet derivative is then a matrix operator with elements that are merely the partial derivatives of $\mathcal{M}$ with respect to each ρ_p. The element corresponding to θ_{meas} and ρ_p is then

$$\mathcal{M}'_p = \partial\mathcal{M}/\partial\rho_p = \sum_{m=-\infty}^{\infty} i^m \partial b_m/\partial\rho_p \exp(im\theta_{\text{meas}}). \tag{31}$$

From (17), the term $\partial b_m/\partial\rho_p$ is conveniently expressed as

$$\partial b_m/\partial\rho_p = \sum_{n=-\infty}^{\infty} (J'_{mnp} f_n + J_{mn} f'_{np}). \tag{32}$$

Note that the primes in the concise notation J'_{mnp} and f'_{np} do not strictly denote differentiation. In fact, from (18),

$$\begin{aligned} J'_{mnp} &= k \int_0^{2\pi} J'_m(k\rho(\phi)) \frac{\partial\rho}{\partial\rho_p}(\phi) \exp(i(n-m)\phi)\, d\phi \\ &= k \int_0^{2\pi} J'_m(k\rho(\phi)) \varrho_p(\phi) \exp(i(n-m)\phi)\, d\phi. \end{aligned} \tag{33}$$

The remaining term, yet to be defined, in (32) is f'_{np}. This can be computed by noting that (14), when differentiated and rearranged reads, for each $p = -\infty, \ldots, \infty$,

$$\sum_{n=-\infty}^{\infty} H_{mn} f'_{np} = -\sum_{n=-\infty}^{\infty} H'_{mnp} f_n, \quad m = -\infty, \ldots, \infty, \tag{34}$$

where H'_{mnp} is defined similarly to J'_{mnp}, that is, (33) with J_m replaced by $C_m H_m^{(2)}$. Note that, regardless of the value of p, the linear system of equations (34) involves the same linear operator as (14). Since only the right hand sides differ, the set $\{f_{np}\}$ can be calculated with little extra work once the inverse operator has been computed to solve (14).

Efforts to efficiently compute J'_{mnp} and H'_{mnp} influence the choice of $\{\varrho_p(\phi)\}$, as demonstrated in the next section.

It is worth noting that, in practical situations, accurate measurement of the phase of ψ_s may be difficult [20] so that a more realistic measurement is of intensity,

$$\mathcal{U}(\rho) = \mathcal{M}(\rho)\mathcal{M}^*(\rho). \tag{35}$$

This alternative form of direct mapping poses no serious complication to the procedures developed above, because each component of the Fréchet derivative of $\mathcal{U}$ is trivial to evaluate,

$$\partial \mathcal{U}/\partial \rho_p = 2\mathrm{Re}(\mathcal{M}^* \partial \mathcal{M}/\partial \rho_p), \tag{36}$$

once $\mathcal{M}'$ has been computed in (31).

Reference to (24) reveals that, when the Neumann boundary condition applies, (33) should be replaced by

$$\begin{aligned} J'_{mnp} \;=\; & -\int_0^{2\pi} \Big\{ \Big[kJ'_m(k\rho(\phi)) + k^2\rho(\phi)J''_m(k\rho(\phi)) + ikm\rho'(\phi) \Big] \varrho_p(\phi) \\ & + imJ_m(k\rho(\phi))\varrho'_p(\phi) \Big\} \exp(i(n-m)\phi)\, d\phi, \end{aligned} \tag{37}$$

the double prime denoting second derivatives, with the expression for H'_{mnp} having J_m replaced by $C_m H_m^{(2)}$.

5 NUMERICAL IMPLEMENTATION

Computation of the $\mathcal{M}'$ entails summation of the expressions in (31) and (32), and solution of the system of equations (34) for all p. Note that $\{J_{mn}\}$ and $\{f_n\}$ can be retrieved from the solution of the direct problem, and require no further computation. It remains to specify particular basis functions $\{\varrho_p(\phi)\}$ that ease the effort in solving (27).

The major component of the additional computational effort associated with the Newton-Kantorovich method is in the evaluation of J'_{mnp} and H'_{mnp} at each iteration. Inspection of (33) reveals that, in order to compute $\{J'_{mnp}\}$ and $\{H'_{mnp}\}$ efficiently, is is desirable to use complex exponential functions for $\{\varrho_p(\phi)\}$. A first choice for the representation of $\rho(\phi)$ might then be

$$\rho(\phi) = \rho_0 + \sum_{p=1}^{\infty} (\rho_p^{(+)} \exp(ip\phi) + \rho_p^{(-)} \exp(-ip\phi)), \tag{38}$$

where the constraints that ρ_0 be real and that $\rho_p^{(+)}$ and $\rho_p^{(-)}$ form complex conjugate pairs must be imposed to ensure that $\rho(\phi)$ is real-valued. Note that although J' and H' depend on three indices, the simple addition theorem for complex exponential functions can be exploited to combine the indices p and n into a single index $q = p + n$. Thus, the complex exponential representation provides for efficient evaluation of J'_{mnp} and H'_{mnp}, by necessitating evaluation of integrals only of the form

$$H'_{mq} = kC_m \int_0^{2\pi} H_m^{(2)}(k\rho(\phi)) \exp(i(q-m)\phi)\, d\phi, \quad m, q = -\infty, \ldots, \infty. \tag{39}$$

The complex representation (38) is, however, inconvenient when solving (27), where the complex conjugacy requirements must be incorporated as constraint equations. Since ρ is a real function of ϕ, it seems natural to choose a real Fourier series representation

$$\rho(\phi) = \rho_0 + \sum_{p=1}^{\infty} (\rho_p^{(c)} \cos p\phi + \rho_p^{(s)} \sin p\phi), \tag{40}$$

where $\{\rho_p^{(c)}, \rho_p^{(s)}\}$ are real numbers. Fortunately, the advantages of both representations can be preserved by noting the simple transformations between the coefficients :

$$\rho_p^{(c)} = \rho_p^{(+)} + \rho_p^{(-)}, \qquad \rho_p^{(s)} = i(\rho_p^{(+)} - \rho_p^{(-)}), \tag{41}$$

$$\rho_p^{(+)} = (\rho_p^{(c)} - i\rho_p^{(s)})/2, \qquad \rho_p^{(-)} = (\rho_p^{(c)} + i\rho_p^{(s)})/2. \tag{42}$$

Consequently, the partial derivatives with respect to these quantities transform as

$$\partial/\partial\rho_p^{(c)} = \partial/\partial\rho_p^{(+)} + \partial/\partial\rho_p^{(-)}, \qquad \partial/\partial\rho_p^{(s)} = i(\partial/\partial\rho_p^{(+)} - \partial/\partial\rho_p^{(-)}), \quad (43)$$

$$\partial/\partial\rho_p^{(+)} = (\partial/\partial\rho_p^{(c)} - i\partial/\partial\rho_p^{(s)})/2, \qquad \partial/\partial\rho_p^{(-)} = (\partial/\partial\rho_p^{(c)} + i\partial/\partial\rho_p^{(s)})/2. \quad (44)$$

Hence a most efficient procedure for computation of $\mathcal{M}'$ is to first associate $\{\rho_p\}$ with $\{\rho_p^{(+)}, \rho_p^{(-)}\}$, introduced in (38), throughout §4 so that $\mathcal{M}'$ is computed in terms of the the complex exponential basis representation. The coordinate transformations (43) are then invoked to yield $\mathcal{M}'$ with respect to the set $\{\rho_p\} = \{\rho_p^{(c)}, \rho_p^{(s)}\}$, introduced in the real Fourier series representation (40), to be employed in solving (27).

It is instructive to think of (12), (14), (16), and (17) as various summations in terms of coefficients of basis functions. In any computer implementation of the algorithm, finite dimensional spanning sets must be extracted from such bases. While it would be desirable to have analytical means by which to estimate the magnitudes of these dimensions required to achieve a specific accuracy, the former are invariably chosen by computational experiment, that is by the unimaginative procedure of increasing the dimension of the spanning sets until convergence is manifest.

Utilisation of finite dimensional subspaces ensures that the linear operators, H and J, used for computation of both of the operators $\mathcal{M}$ and $\mathcal{M}'$, become finite dimensional matrices. Consequently, all infinite summations in the antecedent are truncated to a finite number of terms, so that H_{mn} and H'_{mnp} have indices m, n, and p varying from, say, $-N_U$ to N_U, $-N_f$ to N_f, and $-N_\rho$ to N_ρ respectively. To determine $\{f_n\}$ in (14), we require $N_U \geq N_f$. For J_{mn} and J'_{mnp}, m meed only vary between $-N_b$ and N_b where $N_b < N_U$. A finite number of measurements $N_\mathcal{M}$ is also postulated so that $\mathcal{M}$ maps from $\mathbf{R}^{2N_\rho+1}$ into $\mathbf{R}^{N_\mathcal{M}}$. Clearly, in order to pose a meaningful inverse problem, we require $N_\mathcal{M} > 2N_\rho$.

We suggest that solution of the linear systems defined by the matrix operator H, namely (14) and (34), and the Newton-Kantorovich update equation (27) should be performed by the use of the singular-valued decomposition (SVD) algorithm [21]. Although this algorithm is more computationally expensive than, say, Gaussian elimination, its capacity to solve over-determined systems and to provide accurate estimation of the vector 2-norm condition number of a matrix [22, Chapter 1] justified its use. It should also be noted that the effort expended computing the SVD is only a small part of that required in our complete implementation (which is itself comparable if not better than those of others [9,11]), the greatest portion being the evaluation of H, J, H' and J' by numerical quadrature.

Remember that solution of (34) does not require further matrix inversion because H^{-1} is effectively available from the direct problem, in the form of the SVD of H. By employing the computational procedure described in the paragraph following (43), evaluation of the $\mathcal{M}'$ only entails evaluation of H' and J', and multiplication of previously computed quantities.

The major component of the additional computation effort required to obtain an update is in the evaluation of J' and H' at each iteration, which is comparable to solving the direct problem. However, these quantities are used repeatedly for different components of $\mathcal{M}'$. Thus, unlike a previous method [11], our scheme provides an efficient calculation of $\mathcal{M}'$ as the number of terms used to represent the obstacle is increased.

Note also, that computation of $\mathcal{M}'$ involves quantities that can, in principle, be evaluated to arbitrary accuracy. Numerical inaccuracies are inevitable, but the numerical conditioning of the evaluation of $\mathcal{M}'$ is governed by the conditioning of H,

and so is equal to that of solving the direct problem. Any noise in the data for the inverse problem has no effect on the computation of $\mathcal{M}'$, but of course affects solutions of (27).

Software to solve the direct and inverse problems was programmed on a Digital MicroVAX-II in standard FORTRAN-77 using single precision floating point real numbers with $\epsilon = 1.0e-7$, where ϵ is the smallest (positive) number such that $(1.0+\epsilon)$ produces a result greater 1.0. The integrals involved in computation of H, J, H' and J' were all evaluated to a relative accuracy of at least $1.0e-4$ by an adaptive Simpson algorithm.

Numerical tests on the portion of the program that solved direct problems showed good agreement with the analytical results for cylinders of constant radii, and for other simple geometries described in [17, §3.2.1]. The operator H was truncated to a square matrix of rank N_f. Choosing $N_U = N_f = 15$ and $N_b = 10$, was sufficient to satisfy the forward scattering theorem [17, page 7] to a relative accuracy of at least 1.0e-3 each time a direct problem was solved.

The data for the inverse problem was obtained by solving the direct problem for an obstacle described in dimensionless terms by

$$k\rho(\phi) = 2 + 0.5\cos(4\phi), \tag{45}$$

that is probed by an incident field $\exp(-ikx)$ (see Figure 1). For the inverse problem, N_ρ was set to 4 and $N_{\mathcal{M}}$ to 9. The measurement angles in the far-field were evenly spaced by $2\pi/9$ radians and included $\theta = 0$. The real and imaginary parts of the scattered field were treated as separate equations, so that (27) became an over-determined system of 18 equations in $2N_\rho + 1 = 9$ unknowns. Such over-determination is essential in practical situations where measurement noise is inevitably present. Since our principal objective was to demonstrate the validity of our Fréchet derivative, no noise was added to the measurements. The Newton-Kantorovich algorithm can be modified to account for noise by replacing (25) by an equivalent optimisation problem with appropriate regularising constraints. One way of incorporating the constraints is by utilizing the a penalty function method [11]. The Fréchet derivative evaluated according to the procedures described in §§4 and 5 can be employed directly in the penalty function method, as described in [11], because the standard Gauss-Newton approximation to the Hessian matrix is utilised.

The initial estimate of the obstacle was chosen arbitrarily to be the circle $kr = 2$, shown as the dashed curve in Figure 1. The final reconstruction is visually indistinguishable from the true obstacle. Results of 4 iterations of the Newton-Kantorovich method are shown in Table 1. The conditioning of the direct problem and of the update problem (27) solved at each iteration is listed, as well as the error in the estimate of the obstacle measured in the maximum norm. The error is seen to exhibit quadratic convergence, demonstrating the effectiveness of our algorithm. The typical CPU time to perform one iteration was two minutes.

6 DISCUSSION

We have demonstrated that the inverse problem for an impenetrable obstacle can be formulated as a non-linear operator equation which can be solved by the Newton-Kantorovich algorithm. The use of the null-field method to solve the direct problem leads to a tractable update problem and quadratic convergence can be obtained.

Table 1 show that the conditioning of the inverse problem, as defined by solution of (27), is such that the particular problem considered here with the aforemen-

tioned truncation values (see §5) is not, in fact, ill-conditioned. The effect of noise on the data for this particular problem will not, therefore, be very significant.

It should be noted that formulations such as ours and others [11] that employ the null-field method to define $\mathcal{M}$ implicitly assume that the origin of coordinates can be positioned at a point within Υ_-. Furthermore, the effect of a shift of origin is to alter the phase of the scattered field whilst preserving its magnitude, so that the absolute position of the obstacle can only be determined when the phase of the scattered field is known accurately. The problem of positioning the origin disappears when only the intensity of the scattered field is given. This is because an equivalence class of scatterers exists, consisting of translated versions of the true scatterer, whose members all produce the same data. Consequently, when only the intensity of the scattered field is measured, only the shape of the obstacle can be deduced, and the experimenter is always free to select a representation in which the origin is within Υ_-.

We are also currently implementing our method for problems in which the Neumann boundary condition applies.

7 ACKNOWLEDGEMENTS

We thank our colleagues R. H. T. Bates, T. S. Yeo, and T. J. Connolly for many helpful discussions. One of us (DGHT) acknowledges the award of a Postgraduate Research Scholarship from the New Zealand University Grants Committee and an Australian Guaranteed Corporation (NZ) Ltd. Young Achievers Award 1987.

References

[1] Baltes H. P. editor 1980 *Inverse Scattering Problems in Optics* (Berlin: Springer-Verlag). Topics in Current Physics volume 20.

[2] Berkhout A. J. 1986 The seismic method in the search for oil and gas: current techniques and future developments *Proceedings IEEE* **74** 1133–1159.

[3] Boerner W. M., Jordan A. K. and Kay I. W. editors 1981 Special issue on inverse methods in electromagnetics *IEEE Transactions on Antennas and Propagation* **AP-29** 185–417.

[4] Sabatier P. C. 1983 Theoretical considerations for inverse scattering *Radio Science* **18** 1–18.

[5] Colton D. 1984 The inverse scattering problem for time-harmonic acoustic waves *SIAM Review* **26** 323–350.

[6] Bates R. H. T., Garden K. L. and Peters T. M. 1983 Overview of computerized tomography with emphasis on future developments *Proceedings IEEE* **71** 356–372. Special Issue on Computerized Tomography.

[7] Devaney A. J. 1984 Geophysical diffraction tomography *IEEE Transactions on Geoscience and Remote Sensing* **GE-22** 3–13.

[8] Connolly T. J., Wall D. J. N. and Bates R. H. T. 1985 Inverse problems and the Newton-Kantorovich method. In Devaney A. J. and Bates R. H. T. editors, *Inverse Optics II* 30–34 Proceedings of SPIE volume 558.

[9] Roger A. 1981 Newton-Kantorovitch algorithm applied to an electromagnetic inverse problem *IEEE Transactions on Antennas and Propagation* **AP-29** 232–238.

[10] Wall D. J. N., Yeo T. S. and Bates R. H. T. 1985 Inverse scattering and the null field method. In Devaney A. J. and Bates R. H. T. editors, *Inverse Optics II* 25–29 Proceedings of SPIE volume 558.

[11] Kristensson G. and Vogel C. R. 1986 Inverse problems for acoustic waves using the penalised likelihood method *Inverse Problems* **2** 461–479.

[12] Bates R. H. T. and Wall D. J. N. 1977 Null field approach to scalar diffraction: I,II,III *Philosophical Transactions of the Royal Society of London, Series A: Mathematical and Physical Sciences* **A287** 45–114.

[13] Müller C. 1969 *Foundations of the Mathematical Theory of Electromagnetic Waves* (New York: Springer-Verlag).

[14] Morse P. M. and Feshbach H. 1953 *Methods of Theoretical Physics* (New York: McGraw-Hill). Volume 1 : Chapters 1 to 8; volume 2 : Chapters 9 to 13.

[15] Watson G. N. 1966 *A Treatise on the Theory of Bessel Functions* (Cambridge: Cambridge University Press) second edition.

[16] Wall D. J. N. 1980 Methods of overcoming numerical instabilities associated with the T-matrix method. In Varadan V. K. and Varadan V. V. editors, *Acoustic, Electromagnetic and Elastic Wave Scattering — Focus on the T-matrix approach* 269–286 (New York: Pergamon Press).

[17] Bowman J. J., Senior T. B. A. and Uslenghi P. L. E. editors 1969 *Electromagnetic and Acoustic Scattering by Simple Shapes* (Amsterdam: North-Holland).

[18] Rall L. B. 1969 *Computational Solution of Nonlinear Operator Equations* (New York: Wiley).

[19] Dennis J. E. and Schnabel R. B. 1969 *Numerical Methods for Unconstrained Optimization and Nonlinear Equations* (Engelwood Cliffs, New Jersey: Prentice-Hall).

[20] Bates R. H. T. and Tan D. G. H. 1985 Towards reconstructing phases of inverse-scattering signals *Journal of the Optical Society of America A* **2** 2013–2018.

[21] Businger P. A. and Golub G. H. 1969 Algorithm 358: Singular value decomposition of a complex matrix *Comm. Assoc. Comp. Mach.* **12** 564–565.

[22] Golub G. H. and Van Loan C. F. 1983 *Matrix Computations* (Oxford: North Oxford Academic).

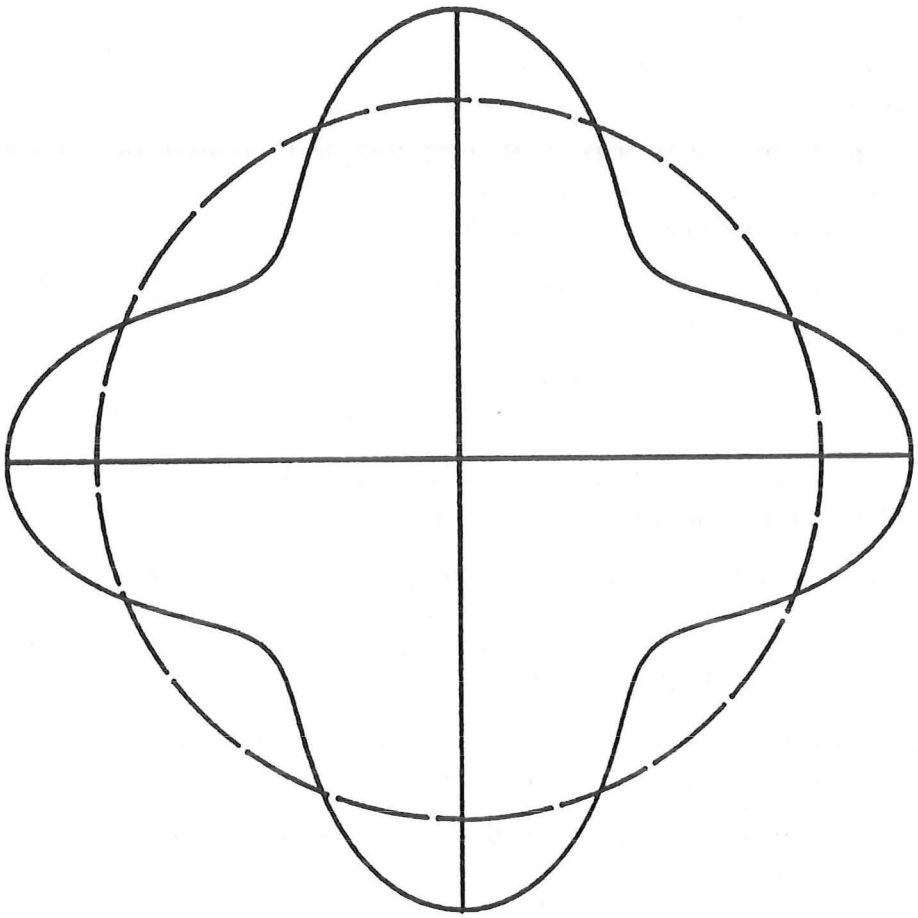

Figure 1: The scattering geometry. Solid line is the scattering obstacle $k\rho(\phi) = 2 + 0.5\cos(4\phi)$ to be reconstructed. Dashed circle is the initial estimate $k\rho = 2$. The final reconstruction is indistinguishable from the true obstacle and so is not displayed.

Iteration	Direct problem condition number	Inverse problem condition number	L_∞ error after update
0			0.5
1	1.	25.446	0.130
2	10.397	38.910	0.0239
3	86.956	40.000	0.000143
4	69.930	38.760	0.0000056

Table 1: Results of Newton-Kantorovich method.

6-C ALGORITHMIC IMPLEMENTATION OF GLOBAL SOLUTION TO SCALAR INVERSE SCATTERING PROBLEM

This appendix is a preprint of a paper (authors: D. G. H. Tan, R. D. Murch, and R. H. T. Bates) submitted (February 1988) to *Inverse Problems*.

Abstract

A numerical solution to the inverse problem for the Helmholtz equation with spatially varying wavenumber is presented. The algorithm is based on a global solution to the inverse problem which expresses the field (within the inhomogeneous part of space from which the wave motion is scattered) in basis functions that explicitly incorporate the scattering data. The numerical solution employs an eigenvalue search together with a Newton-Kantorovich algorithm. While the algorithm is formulated in 3 space dimensions, numerical examples are presented only for fields and inhomogeneities that vary in 2 dimensions (thereby conserving computational effort without transforming away the essential difficulties, as would be the case if we simplified to 1 space dimension). Convergence properties of the algorithm are illustrated.

1 Introduction

We introduce a technique for generating numerical solutions to inverse problems involving the Helmholtz equation with spatially varying wave number. The algorithm is based on the global formalism which we introduced some time ago [1] and discussed further recently [2]. Although numerical methods have been previously devised, for iterating towards solutions to inverse scattering problems by solving successions of direct problems [3,4,5], our approach seems to be the only one so far proposed that expresses the wave motion in basis functions that are uniquely related to the inverse scattering data. We emphasise the very significant differences, as far as inverse scattering problems are concerned, between wave motions that do and do not exhibit refraction. It is well known how to handle the latter, which are characterised by the Schrödinger and plasma wave equations [6,7]. The point is that the refractive index is understood to be spatially invariant for these two equations.

Necessary preliminaries, including a formal statement of the inverse scattering problem, are introduced in Section 2. We set down our formalism in Section 3, while in Section 4 we describe our new algorithm. Section 5 presents numerical results. In Section 6 we assess the significance of our results and discuss what further studies are needed.

Our analytical derivations are for three space dimensions. For computational economy we present numerical solutions only for two space dimensions, which does not reduce the significance of our algorithm, because it is only for a single space dimension that the Helmholtz wave equation with a spatially varying wave number can be reduced to an equation with a spatially constant wave number [8,9].

2 Preliminaries

An arbitrary point P in three-dimensional space Υ is labelled with the spherical polar co-ordinates (r,ϕ,θ) (see figure 1). Space is partitioned according to

$$\Upsilon = \Upsilon_- \cup \Upsilon_{-+} \cup S \cup \Upsilon_{+-} \cup \Upsilon_+, \tag{1}$$

where Υ_- is the sphere $r < a_-$, the spherical annuli $a_- \leq r < a$ and $a < r \leq a_+$ are denoted Υ_{-+} and Υ_{+-} respectively, S is the spherical surface $r = a$, and Υ_+ is all of space for which $r > a_+$. We specify $(a_+ - a)$ and $(a - a_-)$ to be finite.

We consider scalar wave motion which is everywhere linear, implying that it can be examined equally as well in the frequency (i.e. wave number) domain as in the time domain. We prefer the former and write the wave function $\psi = \psi(r, \phi, \theta, k)$ as a function of space and the free-space wave number k. This wave function is required to satisfy

$$\nabla^2\psi + k^2\nu^2\psi = 0, \tag{2}$$

throughout all space, apart from Υ_0 wherein reside the sources of the wave motion. As indicated in figure 1, $\Upsilon_0 \subset \Upsilon_+$. We specify space to be free within $\Upsilon_{-+} \cup S \cup \Upsilon_{+-}$ and the part of Υ_+ not containing Υ_0, and homogeneous throughout Υ_0, implying that the refractive index $\nu = \nu(r, \phi, \theta)$ satisfies

$$\nu(r, \phi, \theta) = 1 \quad \text{for} \quad r \geq a_-. \tag{3}$$

It is worth recalling [2] that, provided ν is allowed to vary with k, quite general linear wave motion can be reduced to the canonical form characterised by (2). For

convenience, we take ν to be independent of k. In fact, as inspection of Sections 3 and 4 reveals, our method of solution is unaffected by whether or not ν depends upon k.

The previous paragraph suggests that we have developed an approach of considerable generality. While this is potentially correct, we must point out that considerable refinement of our numerical techniques is needed before the full power of the approach can be realised. In particular we must devise appropriate regularisation techniques in order to be able to deal with scatterers having dimensions of many wavelengths.

When we specialise to two space dimensions we set $\theta = \pi/2$ and replace r by ρ, so that the co-ordinates of P are (ρ, ϕ). Also $\psi = \psi(\rho, \phi, k)$ and $\nu = \nu(\rho, \phi)$.

The inverse scattering problem is here posed as : given the sources of the wave motion within Υ_0, knowing ψ throughout Υ_{+-}, determine ν for $r < a$.

Because $(a_+ - a)$ is finite, the given information implies that the field scattered from the spatially varying refractive index can be immediately separated from the field (due to the sources) incident upon this varying refractive index. Furthermore, both fields can be taken as known throughout Υ_{+-}. Because $(a - a_-)$ is also finite, the boundary conditions on S are

$$\psi(a^-, \phi, \theta, k) = \psi(a^+, \phi, \theta, k) \quad \text{and} \quad \partial_r \psi(a^-, \phi, \theta, k) = \partial_r \psi(a^+, \phi, \theta, k), \tag{4}$$

where the notation $a^{\pm}$ implies

$$\lim_{\epsilon \to 0} a \pm \epsilon = a^{\pm}, \tag{5}$$

with ϵ being real and positive.

3 Formal solution

Within Υ_+ we express the wave function as

$$\psi = \psi_{inc} + \psi_s, \tag{6}$$

where $\psi_{inc} = \psi_{inc}(r, \phi, \theta, k)$ is the incident field emanating from Υ_0, and $\psi_s = \psi_s(r, \phi, \theta, k)$ is the field scattered from the spatially varying refractive index inside S. We consider a set $\{\psi_{inc}^{p,l}(r, \phi, \theta, k); p = -l, -l+1, ..., l; l = 0, 1, ...\}$ of incident fields, the p, l^{th} being chosen to be

$$\psi_{inc}^{p,l}(r, \phi, \theta, k) = j_l(kr) Y_l^p(\phi, \theta) \quad \text{for} \quad P \in \Upsilon_{+-}, \tag{7}$$

where $j_l(\cdot)$ denotes the spherical Bessel function of the first kind of order l, and $Y_l^p(\cdot, \cdot)$ is the p, l^{th} spherical harmonic regular throughout $0 \leq \theta \leq \pi$ and $0 \leq \phi \leq 2\pi$ [10]. We take the implied temporal variation of the wavefunction to be $\exp(i\omega t)$ where $i = \sqrt{-1}$, $\omega = kc$, c=wave speed in free space and t=time, so that the p, l^{th} scattered field $\psi_s^{p,l}$ can be written as

$$\psi_s^{p,l}(r, \phi, \theta, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} A_{m,n}^{p,l}(k) h_n^{(2)}(kr) Y_n^m(\phi, \theta) \quad \text{for} \quad P \in \Upsilon_{+-}, \tag{8}$$

where $A_{m,n}^{p,l}(k)$ is the m, n^{th} expansion coefficient for the p, l^{th} wavefunction $\psi^{p,l}$, and $h_n^{(2)}(\cdot)$ denotes the spherical Hankel function of the second kind of order n. The $A_{m,n}^{p,l}(k)$ constitute the data for the inverse scattering problem because, as stated in Section 2, the wave function is known throughout Υ_{+-}

An appropriate representation of $\psi^{p,l}$ throughout $\Upsilon_{-+} \cup \Upsilon_-$ is

$$\psi^{p,l}(r, \phi, \theta, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \psi_{m,n}^{p,l}(r, k) Y_n^m(\phi, \theta) \quad \text{for} \quad r < a, \tag{9}$$

where the functional forms of $\psi^{p,l}_{m,n}(r,k)$ are, as yet, unspecified. We define the impedances

$$Z^{p,l}_{m,n}(k) = -\frac{\partial_r \psi^{p,l}_{m,n}(a,k)}{\psi^{p,l}_{m,n}(a,k)}. \tag{10}$$

Reference to (4), (7) and (8) then indicates that

$$Z^{p,l}_{m,n}(k) = -k\frac{j_n'(ka)\delta_{p,m}\delta_{l,n} + A^{p,l}_{m,n}h_n^{(2)'}(ka)}{j_n(ka)\delta_{p,m}\delta_{l,n} + A^{p,l}_{m,n}h_n^{(2)}(ka)}, \tag{11}$$

where a dash (or prime) denotes the total derivative, and the Kronecker delta $\delta_{m,n} = 1$ for $m = n$ and 0 for $m \neq n$ (m and n being integers). Note that

$$Z^{p,l}_{m,n}(k) = -k\frac{h_n^{(2)'}(ka)}{h_n^{(2)}(ka)} \quad \text{for} \quad m \neq p \quad \text{and} \quad n \neq l. \tag{12}$$

Since the $\psi^{p,l}_{inc}(r,\phi,\theta,k)$ are all of unit amplitude, in the sense that there is no multiplier of the standard eigenfunctions on the right hand side of (7), it is appropriate to think of the $Z^{p,l}_{p,l}$ and the $A^{p,l}_{m\neq p,n\neq l}(k)$ as the normalised inverse scattered data.

When (9) is substituted into (2), and the orthogonality of the $Y^m_n(\phi,\theta)$ is invoked, the resulting second order differential equations (with respect to r) seem most readily manipulated if we set

$$\psi^{p,l}_{m,n}(r,k) = \sum_{\mu=1}^{\infty} B^{p,l}_{\mu,m,n}(k) j_n(\alpha^{p,l}_{\mu,m,n} r), \tag{13}$$

where $B^{p,l}_{\mu,m,n}(k)$ is the μ, m, n^{th} expansion coefficient for the p, l^{th} wave function. On requiring that each $\alpha^{p,l}_{\mu,m,n} = \alpha^{p,l}_{\mu,m,n}(k)$ satisfy the eigenvalue relation

$$\alpha^{p,l}_{\mu,m,n} j_n'(\alpha^{p,l}_{\mu,m,n}) = Z^{p,l}_{m,n}(k) j_n(\alpha^{p,l}_{\mu,m,n}), \tag{14}$$

we see that (9) automatically accommodates the normalised inverse scattering data. In fact, as we have remarked previously (although in contexts restricted to two space dimensions [1,2]) the $\alpha^{p,l}_{\mu,m,n}(k)$ are uniquely related to the $Z^{p,l}_{m,n}(k)$. This is because, by an obvious extension of the theory of Dini series [11], $\{j_n(\alpha^{p,l}_{\mu,m,n} r); \mu = 1, 2, ...\}$ is found to be an orthogonal set (with weight function r^2) throughout $0 \leq r \leq a$, for any quartet of integers p, l, m and n, with l and n non-negative and $|p| \leq l$ and $|m| \leq n$. Consequently to recover ν for $r < a$ we must infer either explicitly or implicitly, the $B^{p,l}_{\mu,m,n}(k)$ from the $Z^{p,l}_{m,n}(k)$ and $B^{p,l}_{\mu,m,n}(k)$.

On substituting (9) and (13) into (2), and invoking Bessel's equation [11] to eliminate the term $r^{-2}\partial_r(r^2\partial_r j_n(\alpha^{p,l}_{\mu,m,n} r))$, it transpires that

$$\sum_{\tilde{n}=0}^{\infty}\sum_{\tilde{m}=-\tilde{n}}^{\tilde{n}}\sum_{\tilde{\mu}=1}^{\infty}\left[I^{\tilde{\mu},\tilde{m},\tilde{n},p,l}_{\mu,m,n} - \delta_{\mu,\tilde{\mu}}(\alpha^{p,l}_{\mu,m,n})^2 N^{p,l}_{\mu,m,n}\right] B^{p,l}_{\tilde{\mu},\tilde{m},\tilde{n}}(k) = 0 \quad \text{for} \quad r < a, \tag{15}$$

where

$$\begin{aligned} I^{\tilde{\mu},\tilde{m},\tilde{n},p,l}_{\mu,m,n} = \int_0^a \int_0^{2\pi} \int_0^{\pi} \; & k^2\nu^2(r,\phi,\theta) j_n(\alpha^{p,l}_{\mu,m,n} r) j_{\tilde{n}}(\alpha^{p,l}_{\tilde{\mu},\tilde{m},\tilde{n}} r) \\ & Y^m_n(\phi,\theta) Y^{\tilde{m}}_{\tilde{n}}(\phi,\theta) r^2 \sin\theta \, d\theta \, d\phi \, dr, \end{aligned} \tag{16}$$

and the normalisation constant $N^{p,l}_{\mu,m,n}(k)$ is defined by

$$N^{p,l}_{\mu,m,n}(k) = \int_0^a \int_0^{2\pi} \int_0^{\pi} \left[k j_n(\alpha^{p,l}_{\mu,m,n} r) Y^m_n(\phi,\theta) \right]^2 r^2 \sin\theta \, d\theta \, d\phi \, dr, \qquad (17)$$

We see that (15) represents a triply infinite system of linear algebraic equations for the $B^{p,l}_{\mu,m,n}(k)$. A non-trivial solution exists only if the determinant of the coefficients of the $B^{p,l}_{\mu,m,n}(k)$ vanishes. We know of no direct means of utilising (15) to determine ν nor have we been able to devise any direct way of evaluating this "inverse scattering determinant". The following Section, demonstrates, however, that iterative manipulation of the equations derived in this Section permits the form of $\nu(r,\phi,\theta)$ to be estimated.

An intriguing aspect of the above mentioned determinant is that it suggests it may be possible to recover $\nu^2(\rho,\phi,\theta)$ for a single pair of integers p,l provided the normalised data are given for (effectively) all k. Alternatively, it would seem that a single value of k may be sufficient, if data are available for (effectively) all pairs of integers p,l. These remarks are not offered with any rigorous intent, but we think them worth making because they accord with results obtained for wave motions that do not exhibit refraction [12]. Further comments relevant to this are made in Sections 4 and 6.

When the refractive index is spherically symmetric, so that it depends only upon the radial co-ordinate r, one would expect to be able to recover $\nu^2 = \nu^2(r)$ from data depending upon a single parameter. Inspection of (8) indicates that $A^{p,l}_{m,n}(k) = 0$ unless $m = p$ and $n = l$. So, we expect $A^{p,l}_{p,l}(k)$ to represent sufficient data either for a single pair of integers p,l and a range of k or a single value k and a range of pairs of integers p,l. This accords with what has been established for the non-relativistic quantum mechanical inverse scattering problem[12]. Expressions which are to be evaluated numerically necessarily contain finite numbers of parameters. This implies that we would hope to be able to estimate $\nu^2(r)$ usefully given $A^{p,l}_{m,n}(k)$ at a finite number, say $\hat{\mu}$, of values k_j of k, for a single pair of integers p,l. We would also expect to reconstruct $\nu^2(r)$ to the same order of accuracy from $A^{p,l}_{m,n}(k)$ given for only one value of k, a single integer p, and about the same number $\hat{\mu}$ of integers l. These expectations are borne out by the results presented in Section 5 below.

Since a non-spherically symmetric refractive index requires many more parameters to characterise it to the same accuracy than one that it is spherically symmetric, we expect that more data are needed to be able to reconstruct it usefully. This does not necessarily imply however, that we need take more discrete values of k or more pairs of integers p,l. The point is that for a single pair of integers p,l and one value of k, there are many expansion coefficients $A^{p,l}_{m,n}(k)$. So, provided we are given $A^{p,l}_{m,n}(k_j)$ for J values k_j for a single pair of integers p,l and for all pairs of integers m,n for which $|A^{p,l}_{m,n}|$ is significant, we would expect to be able to recover a useful estimate of $\nu^2(r,\phi,\theta)$. This means that we must supplement the inverse scattering determinant with equations relating the $B^{p,l}_{\mu,m,n}(k)$ to the $A^{p,l}_{m,n}(k)$ for $m \neq p$ and $n \neq l$.

4 Algorithmic Implementation of the Formal Solution

All the conceptual and analytical difficulties associated with our approach (and all other exact approaches, for that matter) to the inverse scattering problem are present in two space dimensions, although the required computational effort is of course much reduced by comparison with three dimensions. So, as already intimated in Section 1,

we conserve "computational energy" by operating on the two-dimensional versions of the equations developed in Section 3. Employing the notation introduced in Section 2, we replace (7), (8), (9), (10), (11) and (13) through (17) by

$$\psi^l_{inc}(\rho,\phi,k) = J_l(k\rho)e^{il\phi}, \tag{18}$$

$$\psi^l_s(\rho,\phi,k) = \sum_{m=0}^{\infty} A^l_m(k)H^{(2)}_m(k\rho)e^{im\phi} \quad \text{for} \quad P \in \Upsilon_{+-}, \tag{19}$$

$$\psi^l(\rho,\phi,k) = \sum_{m=-\infty}^{\infty} \psi^l_m(\rho,k)e^{im\phi} \quad \text{for} \quad \rho \le a, \tag{20}$$

$$Z^l_m(k) = -\frac{\partial_\rho \psi^l_m(a,k)}{\psi^l_m(a,k)} = -k\frac{J'_m(ka)\delta_{l,m} + A^l_m H^{(2)'}_m(ka)}{J_m(ka)\delta_{l,m} + A^l_m H^{(2)}_m(ka)}, \tag{21}$$

$$\psi^l_m(\rho,k) = \sum_{\mu=1}^{\infty} B^l_{\mu,m}(k)J_m(\alpha^l_{\mu,m}\rho) \quad \text{for} \quad \rho \le a, \tag{22}$$

$$\alpha^l_{\mu,m}J'_{\mu,m}(\alpha^l_{\mu,m}) = Z^l_m(k)J_m(\alpha^l_{\mu,m}), \tag{23}$$

$$\sum_{\tilde{m}=-\infty}^{\infty}\sum_{\tilde{\mu}=1}^{\infty}\left[I^{\tilde{\mu},\tilde{m},l}_{\mu,m} - \delta_{\mu,\tilde{\mu}}(\alpha^l_{\mu,m})^2 N^l_{\mu,m}\right] B^l_{\tilde{\mu},\tilde{m}}(k) = 0, \tag{24}$$

$$I^{\tilde{\mu},\tilde{m},l}_{\mu,m} = \int_0^a \int_0^{2\pi} k^2\nu^2(\rho,\phi)J_m(\alpha^l_{\mu,m}\rho)J_{\tilde{m}}(\alpha^l_{\tilde{\mu},\tilde{m}}\rho)e^{im\phi}e^{i\tilde{m}\phi}\rho\, d\phi\, d\rho, \tag{25}$$

$$N^l_{\mu,m} = \int_0^a \int_0^{2\pi} \left[kJ_m(\alpha^l_{\mu,m}\rho)e^{im\phi}\right]^2 \rho\, d\phi\, d\rho, \tag{26}$$

where $J_m(\cdot)$ is the cylindrical (i.e. ordinary) Bessel function of the first kind of order m and $H^{(2)}_m(\cdot)$ is the cylindrical Hankel function of the second kind and order 2. Note that, because there is one less dimension, the quartet p, l, m, n of integers introduced in Section 3 reduces here to the pair l, m.

The $Z^l_m(k)$ and the $A^l_m(k)$ constitute the normalised data for the two-dimensional inverse scattering problem. Because we have not been able to find a direct method of evaluating the $I^{\tilde{\mu},\tilde{m}}_{\mu,m}$ we have had to devise an iterative scheme which is introduced below.

The first step in the solution is to solve the eigenvalue equation (23) to obtain $\alpha^l_{\mu,m}$ for all the needed values of l, μ and m. We accomplish this by a Newton root finding algorithm. The eigenvalues are spaced by close to π for large μ, so the initial estimate of $\alpha^l_{\mu+1,m}$ is taken to be $\alpha^l_{\mu,m}+\pi$. In this way the desired number of eigenvalues can be readily determined numerically.

We express the unknown $\nu^2(\rho,\phi)$ in terms of an appropriate set of basis functions. We choose the representation

$$\nu^2(\rho,\phi) = \sum_{n=1}^{N}\sum_{p=-P}^{P} C_{n,p}\, rect_n(\rho)e^{ip\phi}, \tag{27}$$

where the $C_{n,p}$ are expansion coefficients,

$$rect_n(\rho) = \begin{cases} 1 & \text{for } \frac{(n-1)a}{N} < \rho < \frac{na}{N}, \\ 0 & \text{otherwise,} \end{cases} \tag{28}$$

and N and P are finite positive integers. The finiteness of N and P emphasises that our algorithm leads to numerical solutions, which cannot of course be implemented for infinite numbers of basis functions (or for continuous functions either because they have to be quantised in amplitude and spatially sampled, both of which are to be understood in what follows, although we refrain, for expository convenience, from indicating them explicitly). Finite sets of the integers l, m and p must also be chosen.

Once the $\alpha^l_{\mu,m}$ have been determined, it is expedient to compute quantities

$$\mathcal{I}_n^{l,\mu,m,\tilde{\mu},\tilde{m}} = \int_0^a k^2 rect_n(\rho) J_m(\alpha^l_{\mu,m}\rho) J_{\tilde{m}}(\alpha^l_{\tilde{\mu},\tilde{m}}\rho)\rho\, d\rho, \tag{29}$$

for later use. We employ an adaptive Simpson algorithm for these computations [13].

Because our method is iterative, it needs an initial estimate of ν in order to start it. We find it convenient to take ν to be a constant equal to the average of ν in Υ_-. This gives us the set $\{C^{(0)}_{n,p}\}$ of initial expansion coefficients. The remainder of this Section indicates how sets, which are improved in the sense that they increasingly approach the set $\{C_{n,p}\}$ for the true refractive index, are obtained at each iteration, $\{C^{(i)}_{n,p}\}$ being the set at the i^{th} iteration.

When (27) is substituted into (25) and (18) through (26) and (29) are invoked, it is seen that

$$\sum_{\tilde{m}=-M}^{M} \sum_{\tilde{\mu}=1}^{\hat{\mu}} \left[\sum_{n=1}^{\infty} C_{n,\tilde{m}-m} \mathcal{I}_n^{l,\mu,m,\tilde{\mu},\tilde{m}} - \delta_{\mu,\tilde{\mu}} (\alpha^l_{\mu,m})^2 N^l_{\mu,m} \right] B^l_{\tilde{\mu},\tilde{m}} = 0, \tag{30}$$

where $\mu \in \{1, 2, ..., \hat{\mu}\}, m \in \{-M, -M+1, ..., M\}$ and $\hat{\mu}$ and M are finite positive integers. We take $l \in \{0, 1, 2, ..., L\}$ with L being a finite positive integer. This gives a set of homogeneous equations for the $B^l_{\mu,m}$. Equations (18), (19), (20) and (22) provide a set of non-homogeneous equations relating the given A^l_m to the $B^l_{\mu,m}$, namely

$$\sum_{\mu=1}^{\infty} B^l_{\mu,m} J_m(\alpha^l_{\mu,m} a) = A^l_m H^{(2)}_m(ka) + \delta_{l,m} J_l(ka). \tag{31}$$

Equations (30) and (31) need to be solved simultaneously. We do this by combining (30) and (31) for all l as a single set of non-homogeneous equations and write them for the i^{th} iteration as

$$\mathrm{K}(\vec{\mathrm{C}}^{(i)})\vec{\mathrm{B}}^{(i)} = \vec{\mathrm{A}}, \tag{32}$$

where the elements of the matrix K and vector $\vec{\mathrm{A}}$ are straightforwardly found from inspection of (30) and (31). It should be noted that the majority of the components of the $\vec{\mathrm{A}}$ in (32) are zero.

At each iteration, (32) is solved in a least squares sense and the residue $\vec{\mathrm{F}}^{(i)}$ of this least squares solution is computed as

$$\vec{\mathrm{F}}^{(i)}(\vec{\mathrm{C}}^{(i)}) = \vec{\mathrm{A}} - \mathrm{K}(\vec{\mathrm{C}}^{(i)})\vec{\mathrm{B}}^{(i)}. \tag{33}$$

The functional dependence on $\vec{\mathrm{C}}^{(i)}$ emphasises that K is evaluated using the current estimate of the refractive index. Next the Fréchet derivative [14] of $\vec{\mathrm{F}}$ is formed by calculating central differences.

An update to $\vec{\mathrm{C}}^{(i)}$ is found by solving (in a least squares sense)

$$\mathrm{F}'^{(i)}(\vec{\mathrm{C}}^{(i)})\delta\vec{\mathrm{C}}^{(i)} = -\vec{\mathrm{F}}^{(i)}(\vec{\mathrm{C}}^{(i)}), \tag{34}$$

where $\mathbf{F}'^{(i)}$ is the Fréchet derivative at iteration (i) and

$$\vec{\mathbf{C}}^{(i+1)} = \vec{\mathbf{C}}^{(i)} + \delta\vec{\mathbf{C}}^{(i)}. \tag{35}$$

We take as a measure of distance between two vectors $\vec{\mathbf{C}}$ and $\vec{\mathbf{D}}$, the infinity norm, i.e. the maximum absolute difference between respective components C_j and D_j :

$$||\vec{\mathbf{C}} - \vec{\mathbf{D}}||_\infty = \max_j |C_j - D_j|. \tag{36}$$

The iterative method is terminated once the distance between $\vec{\mathbf{C}}^{(i+1)}$ and $\vec{\mathbf{C}}^{(i)}$ falls below a preset level (chosen to be .005 for the numerical examples presented in Section 5).

Equations (34) and (35) represents the Newton-Kantorovich method of solving the non-linear operator equation

$$\vec{\mathbf{F}}(\vec{\mathbf{C}}^{(i)}) = \vec{\mathbf{0}}, \tag{37}$$

which is satisfied at the true solution but not elsewhere (locally at least).

This provides us with a numerical scheme which is equivalent to forcing the determinant of the coefficients of $B^l_{\mu,m}$ in (24) to vanish while accommodating all of the normalised scattering data.

5 Numerical Examples

To illustrate the performance of the algorithm several numerical examples are presented here. In order to obtain accurate computer-generated inverse scattering data, we computed the scattering from two-dimensional objects of the type depicted in figure 2. The interior of each object is subdivided by circular boundaries. Each region between such boundaries is of a constant refractive index. Each refractive index is real and non-negative but is otherwise arbitrary. Although only four interior circular regions are shown in figure 2, our software permits as many boundaries as we have so far felt to be necessary. We call the circular boundary of radius a the exterior circular boundary.

The wavefunction in each region of a body (typified by figure 2) was written as a finite eigenfunction expansion (involving Bessel and trigonometric functions) satisfying the Helmholtz equation in the region. The expansion coefficients were evaluated by matching the eigenfunction expansions across the circular boundaries. Both the wave function and its first derivative were required to be continuous across each circular boundary. With the aid of the addition theorems [11], "mode-matching" can be effected across the boundaries [15,16,17] permitting straightforward (although tedious) evaluation of the expansion coefficients. We have found it easy enough to transcribe this procedure into a computational algorithm.

The examples presented in Section 5.1 are included to demonstrate that the algorithm does exhibit numerical convergence, even though the series representation for ν and ψ incorporated into the algorithm are necessarily approximate (because they are unavoidably truncated). The results reported in Section 5.2 confirm that the algorithm can be successfully invoked for asymmetrical scatterers.

The scatterers examined in this Section are all small, in the sense that their dimensions are comparable with the wavelength. We reiterate the point made in the second paragraph following (3) that we need to devise appropriate regularisation techniques before we can hope to handle significantly larger scatterers.

5.1 Circularly Symmetric Examples

The required computational effort is greatly reduced when $\nu(\rho,\phi) = \nu(\rho)$. It then transpires that

$$\psi_m^l(\rho,\phi,k) = \psi_l^l(\rho,k)e^{il\phi}, \tag{38}$$

implying for each l, that $A_m^l(k) \equiv 0$ for $m \neq l$.

To force ν to be circularly symmetric, we chose all the circular boundaries,which subdivide the object, to be concentric. In general there are Q boundaries, with the radius of the q^{th} being a_q and $a_Q = a_-$. We refer to figure 2 to indicate how this can be arranged for $Q = 3$. The refractive index of region (iv) is set equal to that of region (ii), so that region (iv) effectively disappears. The boundaries of regions (i) and (ii) are then made concentric with the circle of radius $a_3 = a_-$. The constant refractive indices of regions (i) through (iii) are denoted by ν_1, ν_3 and ν_3 respectively.

As regards the reconstruction algorithm, the initial estimate within Υ_- of ν was for all trials taken to be a constant (the value 7.0 was chosen) and L was always set to $(N-1)$ with one exception discussed below. The important parameters are the integers N and $\hat{\mu}$, which are the number of terms in the series representations for, respectively, $\nu(\rho)$ and $\psi_l^l(\rho,k)$.

Table 2 summarises the performance of our algorithm on the data supplied by the objects described in Table 1, when $\hat{\mu}$ was taken to be 6, N was varied and ka was set equal to 1 where we have taken a_- to be identical to a, for convenience. The only values of N considered were multiples of Q, implying that the locations of discontinuities of ν coincided with those of the true refractive index. The integer in each entry is the number of iterations performed to fulfil the stopping criterion while the real number is the distance, as defined in Section 4, between the final estimate and the true refractive index.

For the entry $Q = 1$, $N = 2$, convergence to the true solution was not achieved for $L = 1, 2$ but was for $L = 3$.

The value of $\hat{\mu}$ determines the degree of approximation inherent in the representation (28) of ψ within Υ_-. While increasing $\hat{\mu}$ improves the level of approximation possible, numerical ill-conditioning of (32) increases. Consequently, there is an optimum choice for $\hat{\mu}$.

5.2 Asymmetric Examples

Our goal in this subsection is to demonstrate that our approach can actually be effective with asymmetric objects. Because of the very considerable computational effort required to generate the inverse scattering data, we only present results for simple objects.

The simplest type of asymmetric problem is one for which the scatterer is actually symmetric but the solution is formulated on the assumption that the body may be asymmetric. We computed the scattering from a single circular region of radius a (with $ka = 1$) of refractive index 3. We then set $N = 1, P = 1, \hat{\mu} = 3, L = 2, M = 1$ and took the initial estimate of ν to be characterised, with reference to (27), by $C_{1,-1} = 0.2, C_{1,0} = 2.5$ and $C_{1,1} = 0.2$. We found that the maximum error between the actual value of ν and the estimate of ν at the 4^{th} iteration was .01%

We also examined the explicitly asymmetric object shown in figure 3. We calculated the scattering for $ka = 1$, $\nu_1 = 3.0 + \delta\nu$, $\nu_2 = 3.0 - \delta\nu$, $\delta\nu = 0.2$ and $\nu_3 = 3.0$ (refer to caption to figure 3 for the definitions of ν_1, ν_2 and ν_3). On setting $N = 1$, $P = 1$, $\hat{\mu} = 3$, $L = 2$, and $M = 1$, and taking the initial estimate of $\nu(\rho;\phi)$ to

be the constant 2.5, we obtained after 5 iterations the estimate

$$\nu^2(\rho;\phi) = 3 + .08\sin\phi + .002\cos\phi, \tag{39}$$

which indicates that the perturbations (represented by the two circular regions of radius b) of the uniform refractive index (of value 3.0) are positioned very nearly symmetrically about the line $\phi = \pm\pi/2$ (as inspection of figure 3 reveals that they indeed are).

We did not consider ourselves justified to expend the necessary computer time to obtain a significantly more faithful estimate of $\nu(\rho;\phi)$ than that represented by (39). Before attempting to do this we need to perfect regularisation techniques of the kind intimated in Section 2. We nevertheless wish to emphasise that the accuracy of (39) is significant, in the sense that the integrated difference, for $\rho < a$, of the actual refractive index and its estimate is only 20%.

6 Conclusions

An inversion algorithm, based on an expression of the wave function which explicitly incorporates the inverse scattering data, has been formulated for both two-dimensional and three dimensional scatterers and implemented for two-dimensional scatterers.

While numerical results have only been obtained for scatterers whose size (in terms of the wavelength) and structural complexity are modest, we have demonstrated that convergence towards the correct refractive index distribution is possible. We are currently studying how to incorporate an efficient regularisation procedure into our inversion algorithm.

This paper highlights the chief difficulty that lies in the way of obtaining accurate inverse scattering data. This is the problem of devising scatters that are adequately versatile (in the sense that their shapes and refractive index distributions are flexible enough to serve as useful models of objects of interest in scientific and technological applications) and yet do not demand inordinate computational effort (if usefully accurate scattered fields are to be generated). The piecewise circular bodies which we have chosen represent a practical compromise, but we would have preferred to have generated data for bodies having smoothly varying refractive indices. We urge those who concern themselves with inverse scattering algorithms to pay increased attention this particular problem area.

References

[1] Bates R.H.T. Global solution to the scalar inverse scattering problem. *J.Phys.A:Math.Gen*, 8(6):L80–82, August 1975.

[2] Bates R.H.T. Full-wave computed tomography part 1 : fundamental theory. *IEE Proceedings,Pt A*, 131(8):610–615, November 1984.

[3] Johnson S.A. and Tracy M.L. Inverse scattering solutions by a sinc basis, multiple source, moment method – part i : theory. *Ultrasonic Imaging*, 5:361–393, 1983.

[4] Angell T.S., Kleinman R.E., and Roach G.F. An inverse transmission problem for the Helmholtz equation. *Inverse Problems*, 3:149–180, 1987.

[5] Roger A. Newton-Kantorovich algorithm applied to an electromagnetic inverse problem. *IEEE Transactions on Antennas and Propagation*, AP-29(2):232–238, March 1981.

[6] Sabatier P.C., editor. *Applied Inverse Problems*, Springer-Verlag, New York, 1985.

[7] Cheney M. and Rose J.H. Three-dimensional inverse scattering : high-frequency analysis of Newtons Marchenko equation. *J.Math.Phys.*, 26(3):436–439, March 1985.

[8] Brekhovskikh L.M. *Waves in Layered Media.* Academic, New York, second edition, 1980.

[9] Bates R.H.T. and Wall D.J.N. Null field approach to scalar diffraction. *Phil. Trans. R. Soc. Lond. A.*, 287:45–78, September 1977.

[10] Morse P.M. and Feshbach H. *Methods of Theoretical Physics.* McGraw-Hill Book Company, New York, 1953.

[11] Watson G.N. *A Treatise on the Theory of Bessel Functions.* University Press, Cambridge, 1966.

[12] Chadan K. and Sabatier P.C. *Inverse Problems in Quantum Scattering Theory. Texts and Monographs in Physics*, Springer-Verlag, New York, 1977.

[13] Conte S.D. and de Boor C. *Elementary Numerical Analysis - an algorithmic approach.* Wiley and sons, New York, third edition, 1980.

[14] Rall L.B. *Computational Solution of Nonlinear Operator Equations.* Wiley, New York, 1969.

[15] Row R.V. Theoretical and experimental study of electromagnetic scattering by two identical conducting cylinders. *J. Appl. Phys.*, 26:666–675, 1955.

[16] Howarth B.A. Multiple scattering resonances between parallel conducting cylinders. *Can. J.*, 51:2415–2427, 1973.

[17] Mittra R. and Wilton D.R. A numerical approach to the determination of electromagnetic scattering characteristics of perfect conductors. *Proc IEEE*, 57:2064–2065, 1969.

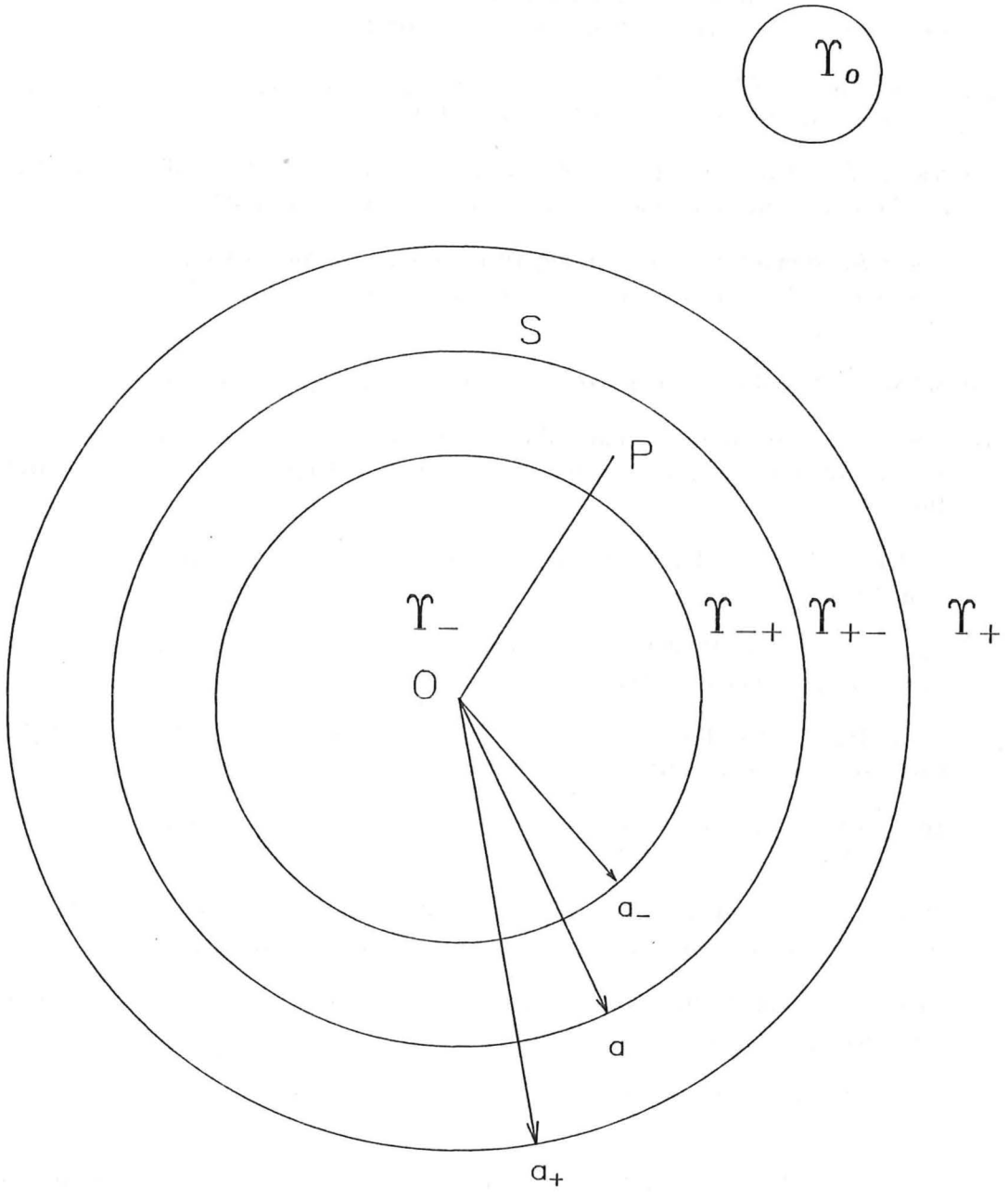

Figure 1: Partitioning of space into spherical regions each centred on the point O.

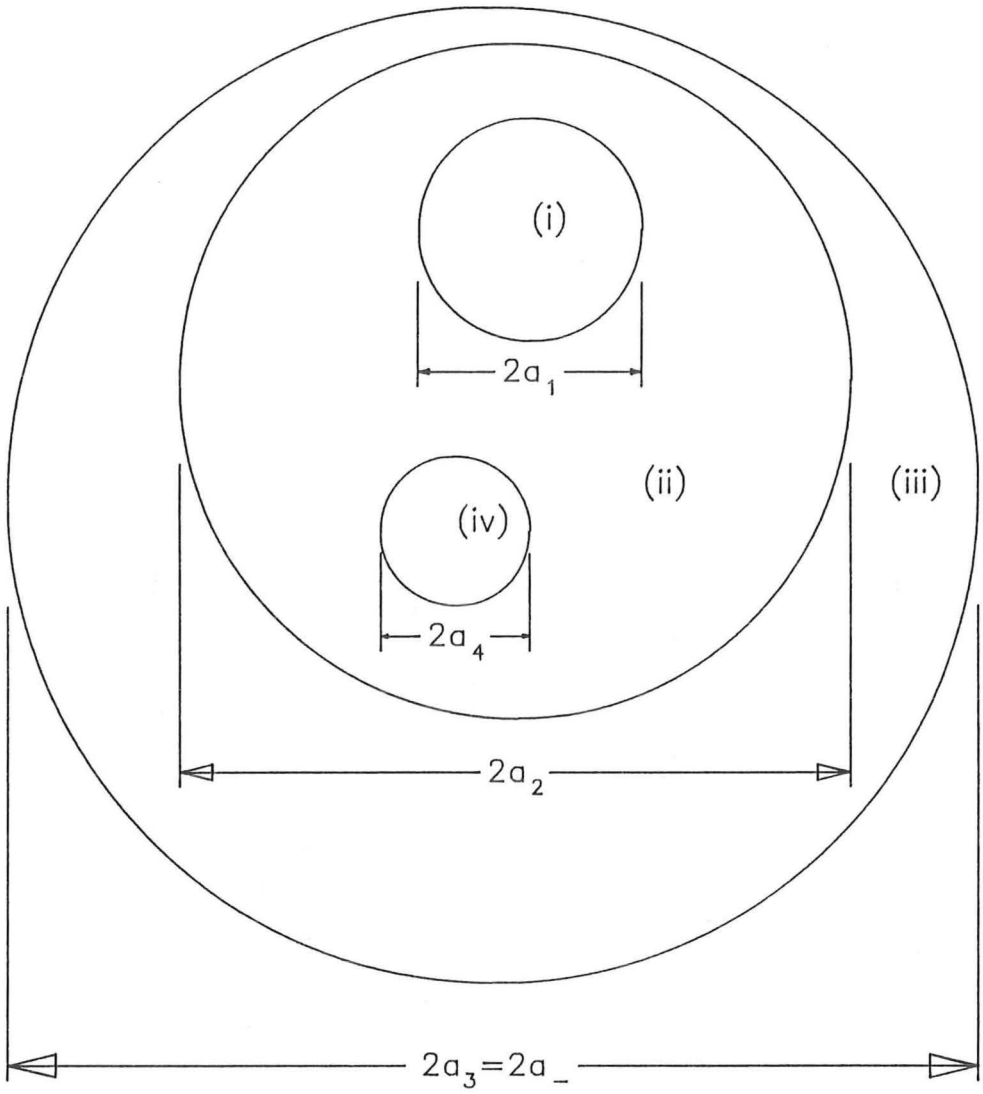

Figure 2: Piecewise two-dimensional object subdivided by circular boundaries. The particular object depicted here has four boundaries. The labelling of the four regions is explained in the text.

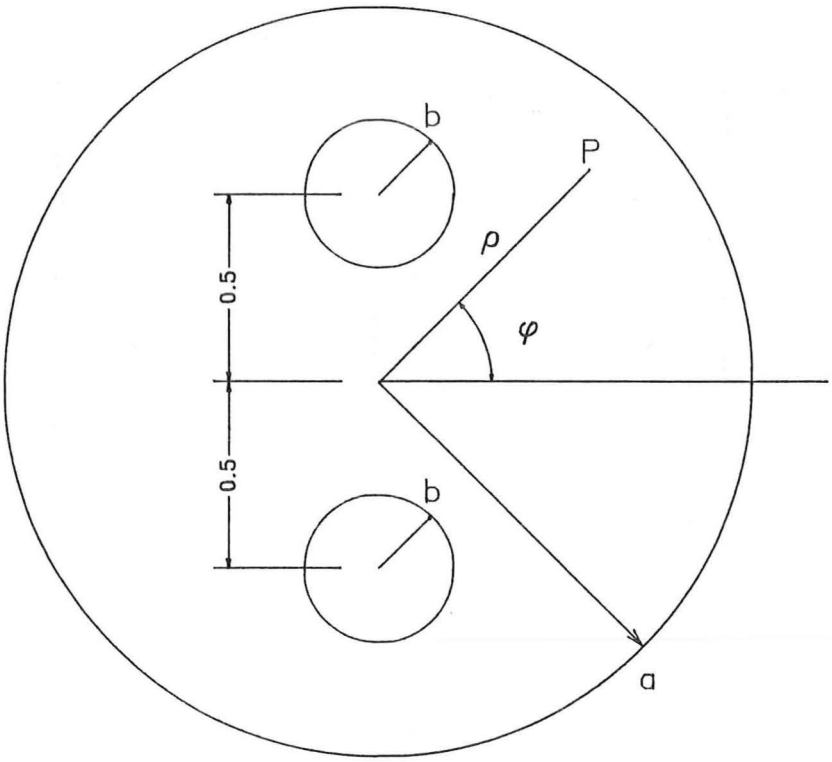

Figure 3: Asymmetric scatterer. P represents an arbitrary point in two dimensions. Refractive indices of upper and lower circular regions of radius b are denoted by ν_1 and ν_2 respectively. Refractive index of remainder of circular region of radius a is denoted by ν_3.

	$\frac{a_q}{\nu_q}$					
Q \ q	1	2	3	4	5	6
1	$\frac{1.0}{7.5}$					
2	$\frac{.5}{6.5}$	$\frac{1.0}{7.5}$				
3	$\frac{.33}{6.5}$	$\frac{.67}{7.0}$	$\frac{1.0}{7.5}$			
4	$\frac{.25}{6.5}$	$\frac{.5}{6.83}$	$\frac{.75}{7.17}$	$\frac{1.0}{7.5}$		
5	$\frac{.2}{6.5}$	$\frac{.4}{6.8}$	$\frac{.6}{7.0}$	$\frac{.8}{7.3}$	$\frac{1.0}{7.5}$	
6	$\frac{.17}{6.5}$	$\frac{.34}{6.7}$	$\frac{.5}{6.9}$	$\frac{.67}{7.1}$	$\frac{.84}{7.3}$	$\frac{1.0}{7.5}$

Table 1: Characteristics of circularly symmetric objects for which results are presented in Section 5.1. Note that a_q and ν_q are respectively, the outer radius and the refractive index of the q^{th} of the Q regions into which each object is subdivided.

Q \ N	1	2	3	4	5	6
1	$\frac{2}{2\text{x}10^{-5}}$	$\frac{4}{2\text{x}10^{-3}}$	$\frac{5}{1\text{x}10^{-2}}$	$\frac{4}{1\text{x}10^{-4}}$	$\frac{3}{1\text{x}10^{-4}}$	$\frac{4}{1\text{x}10^{-3}}$
2		$\frac{2}{2\text{x}10^{-3}}$		$\frac{3}{1\text{x}10^{-3}}$		$\frac{6}{2\text{x}10^{-2}}$
3			$\frac{3}{1\text{x}10^{-3}}$			$\frac{5}{2\text{x}10^{-2}}$
4				$\frac{4}{1\text{x}10^{-2}}$		
5					$\frac{2}{5\text{x}10^{-4}}$	
6						$\frac{9}{5\text{x}10^{-2}}$

Table 2: Summary of the performance of our algorithm on the data supplied by the objects described in Table 1. The integer in each entry is the number of iterations performed to fulfil the stopping criterion while the real number is the distance, as defined in section 4, between the final estimate and the true refractive index.

Chapter 7

CONCLUSIONS AND SUGGESTIONS FOR FUTURE RESEARCH

7.1 CONCLUSIONS

Although inverse problems have been studied for several decades, even centuries, many conceptual and computational difficulties remain. Whereas the theory of direct scattering (i.e. solution of partial differential equations) is well established, the theory of inverse scattering is decidedly tentative by comparison. Tractable numerical inversion schemes are largely limited to problems in which the direct problem is a linear mapping.

The contributions to inverse scattering theory made by this thesis fall into three groups :

(A) It integrates the known direct and inverse approaches into a unified development, of a kind which does not seem to have been attempted previously.

(B) It introduces an extension of diffraction theory, namely, the generalised volume-source formulation developed in Section 3.4.1. The generalised formulation is an improvement over the conventional approach in that it provides scope, through appropriate choice of Green's function, for the incorporation of a particular propagation principle into the description of the diffraction process. Appendix 3-A invokes the generalised formulation to derive (seemingly more systematically than heretofore) a sequence of approximate descriptions of diffraction, potentially applicable in inverse scattering contexts, but falls short of devising inversion algorithms based on the said approximations.

(C) It reports, in Appendices 5-A, 5-B, 5-C, 5-D, 5-E, 6-A, 6-B, and 6-C, eight detailed investigations into particular areas of inverse theory. The most noteworthy aspects of these contributions are as follows :

- Appendices 5-A–5-D demonstrate that various types of extra information (that can often be supplied in practical situations) can, when suitably incorporated into a phase retrieval algorithm, lead to faithful completion of inverse scattering data consisting only of the magnitude of scattered fields.

The ability to successfully perform phase retrieval considerably enhances the potential applicability of many inversion techniques that rely on data consisting of both magnitudes and phases of scattered fields.

- Appendix 5-E provides a quantitative assessment of the effect on images, reconstructed from projections by conventional techniques, of the sampling and resolution of individual projections. The graphical results offer a guide to the amount of detail that can be expected to be discerned by computed tomography systems.
- The model of single photon emission computed tomography developed in Appendix 6-A simultaneously accounts for constant attenuation and variable resolution, making it more sophisticated than previous models which consider each factor in isolation. The model succumbs to lengthy but straightforward analytical manipulation culminating in a relatively simple scheme to improve images formed by applying conventional computed tomography algorithms.
- Appendix 6-B formulates the inverse problem for an impenetrable obstacle (governed by the Dirichlet boundary condition) as a nonlinear operator equation which is then solved by the Newton-Kantorovich algorithm. The invocation of the null-field method to solve the direct problem at each iteration provides an efficient and stable means of evaluating the Fréchet derivative of the direct mapping. Given the numerical complexity of schemes to solve diffraction problems, the exploitation of such structure to reduce computational effort is crucial to the viability of the Newton-Kantorovich algorithm. The formulation can incorporate regularisation, and can be extended to deal with obstacles governed by the Neumann boundary condition. It can also handle situations in which the data supplied are measurements of only the magnitude of the scattered field, making it especially significant in problems for which phase retrieval is not feasible.
- Appendix 6-C presents an algorithmic implementation of a solution to the inverse problem for a general penetrable body that was previously only proposed notionally. A considerable amount of the data is automatically incorporated into the formulation of the direct problem. Whilst the results demonstrate the viability of the technique, efficient means of regularisation and of evaluating the Fréchet derivative will be required before the method can be usefully applied to real-world data.

7.2 SUGGESTIONS FOR FUTURE RESEARCH

The individual problem areas which are most urgently in need of further systematic study, and which offer reasonable hope of yielding to determined assault, are

General properties of the direct problem: The defining relations, and hence existence, uniqueness, and stability properties, of inverse problems stem from the associated direct problems. It is only when the direct mapping is thoroughly understood that reliable interpretation of inverse scattering algorithms is possible.

Analysis of the direct problem determines the types of regularisation appropriate for the corresponding ill-posed inverse problems.

Investigation of the generalised volume-source formulation of diffraction: As concerns direct problems, where the generalised constitutive parameter is given, computationally viable ray-tracing schemes exist for computing the point characteristic function and, hence, the LJWKB Green's function. The generalised volume-source formulation then becomes a linear Fredholm integral equation of the second kind. The numerical viability of both iterative and non-iterative solutions of such an equation are unknown at present. If the generalised volume-source formulation admits perturbation analysis, then it is a candidate for incorporation into a Newton-Kantorovich scheme.

Refinement of phase retrieval algorithms: The theory of zero-sheets suggests that it may soon be possible to perform phase retrieval and deconvolution without prior estimates of the functions involved, as Lane and Bates [1987b] have recently indicated by example. The image processing community has, up to now, found most success with employment of iterative Fienup algorithms. They are simple to implement numerically and the uncertainty in their convergence properties adds to the delight when reconstruction is successful. Any methods of incorporating extra information as constraints within such an algorithm, in such a way that convergence is accelerated, would be beneficial. Of theoretical and practical interest would be a study of the degree to which the amount of extra information imposed can be reduced while still permitting convergence to the true solution to be attained.

Extensions of conventional computed tomography: Computed tomography (CT), based on the projection model of Section 5.4, is sufficiently representative of the physical processes involved in transmission of X-rays through biological tissue to be routinely invoked for medical diagnostic radiology. Advances would accrue from the extension of conventional projection theory to handle situations in which the straight ray model is insufficiently descriptive of the underlying physical process. There are several aspects of CT algorithms that warrant further detailed attention. There is the need for improved allowance for refraction, when the emanations are macroscopic wave motions (e.g. medical ultrasound, sonar, seismic sounding). While this has been partially accomplished for situations in which refractive index variations can be reasonably treated as perturbations to a constant distribution (e.g. as for ultrasonic CT of the female breast), and for situations involving one spatial dimension (e.g. stratified models of the earth used in seismic contexts), one feels that it should be possible to construct a much more widely applicable algorithm. Something else worth pointing out is the rather simplistic nature of the models on which SPECT algorithms are based. It has so far only been practicable to account for attenuation by the body when the attenuation coefficient is assumed constant. Although useful clinical results are obtained thereby, one would hope that variable attenuation (which always occurs in the read-world) could be taken into account by appropriate incorporation of data obtained from a preliminary conventional X-ray CT scan. This possibility has often been mooted of course and tried, but with so little success that very

little definite information concerning it can be found in the literature, although the recent attempt by Heike [1987], which incidentally comes under the Newton-Kantorovich umbrella, may eventually prove to be a useful advance. Similar comments can be made about beam-hardening in X-ray CT (refer to Section 5.4) although it must be emphasised that neglecting it in reconstruction algorithms only results in severe image degradation under restricted sets of the conditions occurring in clinical practice. Nevertheless, it could be worthwhile implementing a convenient means of correcting beam-hardening artefacts, if such could be devised (it would have to be iterative of course).

Speckle Imaging: There is considerable scope for investigation of the performance of ultrasonic speckle imaging under a variety of different schemes for gathering individual speckle images. It has already been established that shift-and-add has the ability to produce faithful images from ensembles of speckle images formed by illuminating the object with plane waves in different narrow frequency bands, all possessing the same direction of incidence. A natural extension is to form speckle images from a variety of incident directions, with frequency either fixed or varied. The extent to which the blurred version of the true image inherent in each speckle image dominates the contamination relies on the ability to usefully assume a simple form for the refractive index of the inhomogeneous medium. One of the most drastic simplifications, which is attractive because it does not require any prior knowledge about the inhomogeneity, is to assume that the inhomogeneity varies negligibly from free-space. If some prior estimate of the generally inhomogeneous refractive index is available, then it seems likely that such information should be usefully incorporated into a more faithful description of the associated direct and inverse problems. An alternative is to estimate the average refractive index throughout the inhomogeneous region. Although theoretical analysis may be feasible, it seems more profitable to undertake experimental investigations with real or simulated (i.e. computer generated) data.

Application of the Newton-Kantorovich method: The Newton-Kantorovich algorithm in its abstract form is a technique of considerable generality. Most significantly, the difficulties of analytical inversion of nonlinear direct mappings are circumvented. It relies, however, on the ability to define the direct problem in terms of an operator whose Fréchet derivative can be readily evaluated numerically or, preferably, determined analytically. The computational effort associated with the Newton-Kantorovich algorithm is sufficiently demanding that the method would find more widespread application if there existed more computationally efficient numerical solutions of direct problems.

Computers can be employed to manipulate large quantities of data at much greater speeds than can people. Any information extracted from such processing of data is only useful if it can be readily interpreted. The ability of the human brain to discern detail from unfamiliar scenes remains unrivalled by automatic computational schemes. For this reason, the discipline of digital image processing is certain to play a significant role in any practical imaging system.

Perhaps the most important lesson to be learned is that the diverse disciplines in which inverse problems arise can benefit from cross-fertilisation of ideas developed

in other applications. Mathematics provides a setting in which the similarities and differences of particular problems can be appreciated. The rigorous aspects of mathematical analysis should not be permitted to overshadow the need to arrive at practical solutions. On the other hand, a thorough understanding of the effectiveness of a particular algorithm can often only be assessed by such an approach. Consequently, there is also a need for continued interaction between theoreticians and experimenters. It is one of the author's aims to be part of such interaction.

References

[Abbott and Thurstone 1979] J. G. Abbott and F. L. Thurstone. Acoustic speckle: Theory and experimental analysis. *Ultrasonic Imaging*, 1(4):303–324, October 1979.

[Abramowitz and Stegun 1965] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions.* Dover Publications, New York, 1965.

[Alvarez and Macovski 1976] R. E. Alvarez and A. Macovski. Energy-selective reconstructions in X-ray computerized tomography. *Physics in Medicine and Biology*, 21(5):733–744, September 1976.

[Andersen 1987] A. H. Andersen. Tomography transform and inverse in geometrical optics. *Journal of the Optical Society of America A*, 4(8):1385–1395, August 1987.

[Andersen and Kak 1984] A. H. Andersen and A. C. Kak. Simultaneous algebraic reconstruction technique (SART): A superior implementation of the ART algorithm. *Ultrasonic Imaging*, 6(1):81–94, January 1984.

[Anderson and Sali 1985] A. P. Anderson and S. Sali. New possibilities for phaseless microwave diagnostics. Part 1: Error reduction techniques. *IEE Proceedings Part H*, H132(5):291–298, August 1985.

[Angell *et al.* 1987] T. S. Angell, R. E. Kleinman, and G. F. Roach. An inverse transmission problem for the Helmholtz equation. *Inverse Problems*, 3(2):149–180, May 1987.

[Arfken 1970] G. Arfken. *Mathematical Methods for Physicists.* Academic Press, New York, international edition, 1970.

[Arsac 1966] J. Arsac. *Fourier Transforms and the Theory of Distributions. Prentice-Hall Applied Mathematics Series*, Prentice-Hall, Englewood Cliffs, New Jersey, 1966. Translated by A. Nussbaum and G. C. Heim.

[Arsenault and Chalasinska-Macukow 1983] H. A. Arsenault and K. Chalasinska-Macukow. The solution to the phase retrieval problem using the sampling theorem. *Optics Communications*, 47(6):380–386, 15 October 1983.

[Assal and Mahmoud 1987] H. M. Assal and S. F. Mahmoud. A new inversion technique for apparent resistivity measurement. *IEEE Transactions on Geoscience and Remote Sensing*, GE-25(1):7–10, January 1987.

[Axel *et al.* 1983] L. Axel, P. H. Arger, and R. A. Zimmerman. Applications of computerized tomography to diagnostic radiology. *Proceedings IEEE*, 71(3):293–297, March 1983. Special Issue on Computerized Tomography.

[Axelsson *et al.* 1987] B. Axelsson, A. Israelsson, and S. A. Larsson. Studies of a technique for attenuation correction in single photon emission computed tomography. *Physics in Medicine and Biology*, 32(6):737–749, June 1987.

[Azimi and Kak 1983] M. Azimi and A. C. Kak. Distortion in diffraction tomography caused by multiple scattering. *IEEE Transactions on Medical Imaging*, MI-2(4):176–195, December 1983.

[Baba *et al.* 1985] N. Baba, S. Isobe, Y. Norimoto, and M. Noguchi. Stellar speckle image reconstruction by the shift-and-add method. *Applied Optics*, 24(10):1403–1405, 15 May 1985.

[Backus and Gilbert 1970] G. Backus and F. Gilbert. Uniqueness in the inversion of inaccurate gross earth data. *Philosophical Transactions of the Royal Society of London, Series A: Mathematical and Physical Sciences*, A266:123–192, 5 March 1970.

[Baker and Copson 1950] B. B. Baker and E. T. Copson. *The Mathematical Theory of Huygens' Principle.* Clarendon, Oxford, second edition, 1950.

[Baltes 1978] H. P. Baltes, editor. *Inverse Source Problems in Optics.* Springer-Verlag, Berlin, 1978. Topics in Current Physics volume 9.

[Baltes 1980] H. P. Baltes, editor. *Inverse Scattering Problems in Optics.* Springer-Verlag, Berlin, 1980. Topics in Current Physics volume 20.

[Barber and Brown 1984] D. Barber and B. Brown. Applied potential tomography. *Journal of Physics E: Scientific Instruments*, 17(9):723–733, September 1984.

[Barcilon 1974] V. Barcilon. On the uniqueness of inverse eigenvalue problems. *Geophysical Journal of the Royal Astronomical Society*, 38:287–298, 1974.

[Barcilon 1976] V. Barcilon. Inverse problem for a vibrating beam. *Journal of Applied Mathematics and Physics*, 27(3):347–358, May 1976. Also known as *Zeitschrift für angewandte Mathematik und Physik.*

[Barcilon 1979] V. Barcilon. On the multiplicity of solutions of the inverse problem for a vibrating beam. *SIAM Journal of Applied Mathematics*, 37(3):605–613, December 1979.

[Barcilon 1982] V. Barcilon. Inverse problem for the vibrating beam in the free-clamped configuration. *Philosophical Transactions of the Royal Society of London, Series A: Mathematical and Physical Sciences*, A304(1483):211–251, February 1982.

[Barcilon 1987] V. Barcilon. Sufficient conditions for the solution of the inverse problem for a vibrating beam. *Inverse Problems*, 3(2):181–193, May 1987.

[Baribaud *et al.* 1985] M. Baribaud, F. Dubois, D. en Méd, R. Floyrac, and S. Wang. Tomographic image reconstitution of objects from multi-incidence microwave exploration. *IEE Proceedings Part H*, 132(5):286–290, August 1985.

[Barrett 1984] H. H. Barrett. The Radon transform and its applications. In E. Wolf, editor, *Progress in Optics volume XXI*, chapter III, pages 217–286, North-Holland Physics, Amsterdam, 1984. See [Wolf 1984].

[Bates 1968] R. H. T. Bates. Modal expansions for electromagnetic scattering from perfectly conducting cylinders of arbitrary cross-section. *Proceedings IEE*, 115(10):1443–1445, October 1968.

[Bates 1974] R. H. T. Bates. New justification for physical optics and the aperture field method. In *Proceedings 20th AGAARD Meeting on Electromagnetic Wave Propagation involving Irregular Surfaces and Inhomogeneous Media (The Hague, The Netherlands)*, pages 36-1–36-7, 1974. AGAARD Conference Publications CPP-144.

[Bates 1975a] R. H. T. Bates. Analytic constraints on electromagnetic field computations. *IEEE Transactions on Microwave Theory and Techniques*, MTT-23(8):605–623, August 1975.

[Bates 1975b] R. H. T. Bates. Global solution to the inverse scattering problem. *Journal of Physics A: Mathematical and General*, 8(8):L80–L82, August 1975.

[Bates 1982] R. H. T. Bates. Astronomical speckle imaging. *Physics Reports*, 90(4):203–297, October 1982.

[Bates 1984] R. H. T. Bates. Full-wave computed tomography. Part 1: Fundamental theory. *IEE Proceedings Part A*, A131(8):610–615, November 1984.

[Bates and Cady 1980] R. H. T. Bates and F. M. Cady. Towards true imaging by wide-band speckle interferometry. *Optics Communications*, 32(3):365–369, March 1980.

[Bates and Davey 1988] R. H. T. Bates and B. L. K. Davey. Towards making shift-and-add a versatile imaging technique. In *Proceedings SPIE 828*, 1988. Paper number 13.

[Bates and Fright 1983] R. H. T. Bates and W. R. Fright. Composite two-dimensional phase restoration procedure. *Journal of the Optical Society of America*, 73(3):358–365, March 1983.

[Bates and Fright 1985] R. H. T. Bates and W. R. Fright. Two-dimensional phase restoration. In J. F. Price, editor, *Fourier Techniques and Applications*, pages 121–148, Plenum Press, New York, 1985.

[Bates and McDonnell 1986] R. H. T. Bates and M. J. McDonnell. *Image Restoration and Reconstruction. Oxford engineering science series volume 16*, Clarendon Press, Oxford, 1986.

[Bates and Millane 1981] R. H. T. Bates and R. P. Millane. Time domain approach to inverse scattering. *IEEE Transactions on Antennas and Propagation*, AP-29(2):359–363, March 1981. Special Issue on Inverse Methods in Electromagnetics.

[Bates and Minard 1984] R. H. T. Bates and R. A. Minard. Compensation for multiple reflection. *IEEE Transactions on Sonics and Ultrasonics*, SU-31(4):330–336, July 1984. Special Issue on Digital Acoustical Imaging.

[Bates and Mnyama 1986] R. H. T. Bates and D. Mnyama. The status of practical Fourier phase retrieval. In P. W. Hawkes, editor, *Advances in Electronics and Electron Physics volume 67*, pages 1–64, Academic Press, Orlando, Florida, 1986.

[Bates and Ng 1972] R. H. T. Bates and F. L. Ng. Polarisation-source formulation of electromagnetism and dielectric-loaded waveguides. *Proceedings IEE*, 119(11):1568–1574, November 1972.

[Bates and Robinson 1981] R. H. T. Bates and B. S. Robinson. Ultrasonic transmission speckle imaging. *Ultrasonic Imaging*, 3(4):378–394, October 1981.

[Bates and Tan 1985a] R. H. T. Bates and D. G. H. Tan. Fourier phase retrieval when the image is complex. In A. J. Devaney and R. H. T. Bates, editors, *Inverse Optics II*, pages 54–59, August 1985. Proceedings SPIE volume 558 (the Society of Photo-Optical Instrumentation Engineers).

[Bates and Tan 1985b] R. H. T. Bates and D. G. H. Tan. Towards reconstructing phases of inverse-scattering signals. *Journal of the Optical Society of America A*, 2(11):2013–2018, November 1985. Special Issue on Inverse Problems in Propagation and Scattering.

[Bates and Tan 1986] R. H. T. Bates and D. G. H. Tan. Phase retrieval for electron diffraction patterns suggesting orientational order without translational symmetry. *Optik*, 73(2):69–73, May 1986.

[Bates and Wall 1976] R. H. T. Bates and D. J. N. Wall. Chandrasekhar transformations improve convergence of computations of scattering from linearly stratified media. *IEEE Transactions on Antennas and Propagation*, AP-24(2):251–253, March 1976.

[Bates and Wall 1977a] R. H. T. Bates and D. J. N. Wall. Null field approach to scalar diffraction: I. General method. *Philosophical Transactions of the Royal Society of London, Series A: Mathematical and Physical Sciences*, A287(1339):45–78, September 1977.

[Bates and Wall 1977b] R. H. T. Bates and D. J. N. Wall. Null field approach to scalar diffraction: II. Approximate methods. *Philosophical Transactions of the Royal Society of London, Series A: Mathematical and Physical Sciences*, A287(1339):79–95, September 1977.

[Bates and Wall 1977c] R. H. T. Bates and D. J. N. Wall. Null field approach to scalar diffraction. III. Inverse methods. *Philosophical Transactions of the Royal Society of London, Series A: Mathematical and Physical Sciences*, A287(1339):97–114, September 1977.

[Bates *et al.* 1976] R. H. T. Bates, W. M. Boerner, and R. G. Dunlop. An extended Rytov approximation and its significance for remote sensing and inverse scattering. *Optics Communications*, 18(4):421–423, September 1976.

[Bates *et al.* 1980] R. H. T. Bates, G. C. McKinnon, and A. D. Seagar. A limitation on sytems for imaging electrical conductivity distribution. *IEEE Transactions on Biomedical Engineering*, BME-27(7):418–420, July 1980.

[Bates *et al.* 1983] R. H. T. Bates, K. L. Garden, and T. M. Peters. Overview of computerized tomography with emphasis on future developments. *Proceedings IEEE*, 71(3):356–372, March 1983. Special Issue on Computerized Tomography.

[Bates *et al.* 1985] R. H. T. Bates, T. S. Yeo, and D. J. N. Wall. Inverse source approach to inverse scattering. In A. J. Devaney and R. H. T. Bates, editors, *Inverse Optics II*, pages 21–24, August 1985. Proceedings SPIE volume 558 (the Society of Photo-Optical Instrumentation Engineers).

[Baum 1976] C. E. Baum. The singularity expansion method. In L. B. Felsen, editor, *Transient Electromagnetic Fields*, chapter 3, pages 129–179, Springer-Verlag, Berlin, 1976. Topics in Applied Physics volume 10.

[Beckmann and Spizzichino 1963] P. Beckmann and A. Spizzichino. *The Scattering of Electromagnetic Waves from Rough Surfaces. International Series of Monographs on Electromagnetic Waves volume 4*, Pergamon Press, Oxford, 1963.

[Bennett and Byer 1986] K. E. Bennett and R. L. Byer. Fan-beam-tomography noise theory. *Journal of the Optical Society of America A*, 3(5):624–633, May 1986.

[Beran and Parrent, Jr. 1974] M. J. Beran and G. B. Parrent, Jr. *Theory of Partial Coherence.* The Society of Photo-Optical Instrumentation Engineers, 1974.

[Beylkin 1985] G. Beylkin. Imaging of discontinuities in the inverse scattering problem by inversion of a causal generalized Radon transform. *Journal of Mathematical Physics*, 26(1):99–108, January 1985.

[Blackledge *et al.* 1987a] J. M. Blackledge, R. E. Burge, K. I. Hopcraft, and R. J. Wombell. Quantitative diffraction tomography: I. Pulsed acoustic fields. *Journal of Physics D : Applied Physics*, 20(1):1–10, January 1987.

[Blackledge *et al.* 1987b] J. M. Blackledge, R. E. Burge, K. I. Hopcraft, and R. J. Wombell. Quantitative diffraction tomography: II. Pulsed elastic waves. *Journal of Physics D : Applied Physics*, 20(1):11–17, January 1987.

[Bleistein and Cohen 1977] N. Bleistein and J. K. Cohen. Non-uniqueness in the inverse source problem in acoustics and electromagnetics. *Journal of Mathematical Physics*, 18(2):194–201, February 1977.

[Bleistein and Cohen 1980] N. Bleistein and J. K. Cohen. Progress on a mathematical inversion technique for non-destructive testing. *Wave Motion*, 2(1):75–82, January 1980.

[Boerner *et al.* 1981] W. M. Boerner, A. K. Jordan, and I. W. Kay, editors. Special issue on inverse methods in electromagnetics. *IEEE Transactions on Antennas and Propagation*, AP-29(2):185–417, March 1981.

[Bond and Reynolds 1987] L. J. Bond and W. N. Reynolds, editors. Special issue on non-destructive testing. *IEE Proceedings Part A*, A134(3):237–306, March 1987.

[Born and Wolf 1970] M. Born and E. Wolf. *Principles of Optics.* Pergamon Press, fourth edition, 1970.

[Bouwkamp 1954] C. J. Bouwkamp. Diffraction theory. *Reports on Progress in Physics*, 17:35–100, 1954.

[Bowman *et al.* 1969] J. J. Bowman, T. B. A. Senior, and P. L. E. Uslenghi, editors. *Electromagnetic and Acoustic Scattering by Simple Shapes.* North-Holland, Amsterdam, 1969.

[Boyse and Keller 1986] W. E. Boyse and J. B. Keller. Inverse elastic scattering in three dimensions. *Journal of the Acoustical Society of America*, 79(2):215–218, February 1986.

[Bracewell 1978] R. N. Bracewell. *The Fourier Transform and its Applications.* McGraw-Hill, New York, second edition, 1978.

[Bracewell and Riddle 1967] R. N. Bracewell and A. C. Riddle. Inversion of fan-beam scans in radio-astronomy. *Astrophysical Journal*, 150(2 Part 1):427–434, November 1967.

[Brekhovskikh 1960] L. M. Brekhovskikh. *Waves in Layered Media.* Academic Press, New York, 1960.

[Brigham 1974] E. O. Brigham. *The Fast Fourier Transform.* Prentice-Hall, Englewood Cliffs, New Jersey, 1974.

[Brooks and Di Chiro 1976a] R. A. Brooks and G. Di Chiro. Beam hardening in X-ray reconstructive tomography. *Physics in Medicine and Biology*, 21(3):390–398, May 1976.

[Brooks and Di Chiro 1976b] R. A. Brooks and G. Di Chiro. Principles of computer assisted tomography (CAT) in radiographic and radioisotopic imaging. *Physics in Medicine and Biology*, 21(5):689–732, September 1976.

[Brown *et al.* 1985] B. H. Brown, D. C. Barber, and A. D. Seagar. Applied potential tomography: Possible clinical applications. *Clinical Physics and Physiological Measurement*, 6(2):109–121, 1985.

[Brown *et al.* 1987] B. Brown, D. Barber, and L. Tarassenko, editors. Special issue on electrical impedance tomography — applied potential tomography. *Clinical Physics and Physiological Measurement*, 8(Supplement A):3–184, 1987. Proceedings of the Commission of the European Communities COMAC-BME Workshop, 2–4 July 1986, Sheffield.

[Bruck and Sodin 1979] Y. M. Bruck and L. G. Sodin. On the ambiguity of the image reconstruction problem. *Optics Communications*, 30(3):304–308, September 1979.

[Budden 1985] K. G. Budden. *The Propagation of Radio Waves: The Theory of Radio Waves of Low Power in the Ionosphere and Magnetosphere.* Cambridge University Press, Cambridge, 1985.

[Budinger and Gullberg 1977] T. F. Budinger and G. T. Gullberg. Transverse section reconstruction of gamma-ray emitting radionuclides in patients. In M. M. Ter-Pogossian *et al.*, editors, *Reconstruction Tomography in Diagnostic Radiology and Nuclear Medicine*, pages 315–342, University Park Press, Baltimore, 1977.

[Budinger *et al.* 1977] T. F. Budinger, S. E. Derenzo, G. T. Gullberg, W. L. Greenberg, and R. H. Huesman. Emission computer assisted tomography with single-photon and positron annihilation photon emitters. *Journal of Computer Assisted Tomography*, 1(1):131–145, 1977.

[Burckhardt 1978] C. B. Burckhardt. Speckle in ultrasound B-mode scans. *IEEE Transactions on Sonics and Ultrasonics*, SU-25(1):1–16, January 1978.

[Byrne and Fiddy 1987] C. L. Byrne and M. A. Fiddy. Estimation of continuous object distributions from limited Fourier magnitude measurements. *Journal of the Optical Society of America A*, 4(1):112–117, January 1987. Special Issue on Signal Recovery.

[Calderone *et al.* 1985] S. Calderone, G. Intilisano, P. G. Mantica, I. Montrosset, and R. Orta. Beam propagation in the ocean: Quasi-particle approach. *IEEE Journal of Oceanic Engineering*, OE-10(3):199–205, July 1985.

[Cannon and Hornung 1986] J. R. Cannon and U. Hornung, editors. *Inverse Problems: Proceedings of Conference Held at the Mathematical Research Institute at Oberwolfach, Black Forest, May 18–24*, Birkhäuser Verlag, Basel, 1986.

[Carter and Wolf 1985] W. H. Carter and E. Wolf. Inverse problem with quasi-homogeneous sources. *Journal of the Optical Society of America A*, 2(11):1994–2000, November 1985. Special Issue on Inverse Problems in Propagation and Scattering.

[Chadan and Sabatier 1977] K. Chadan and P. C. Sabatier. *Inverse Problems in Quantum Scattering Theory. Texts and Monographs in Physics*, Springer-Verlag, New York, 1977.

[Chang 1978] L. T. Chang. A method for attenuation correction in radionuclide computed tomography. *IEEE Transactions on Nuclear Science*, NS-25(1):638–643, February 1978.

[Cheney 1969] E. W. Cheney. *Introduction to Approximation Theory. International Series in Pure and Applied Mathematics*, McGraw-Hill, New York, 1969.

[Cho 1974] Z. H. Cho, editor. Special issue on physical and computational aspects of 3-dimensional image reconstruction. *IEEE Transactions on Nuclear Science*, NS-21(3):1–95, June 1974.

[Cho and Nalcioglu 1984] Z. H. Cho and O. Nalcioglu, editors. Special issue on physics and engineering in nuclear magnetic resonance imaging. *IEEE Transactions on Nuclear Science*, NS-31(4):973–1010, August 1984.

[Cho *et al.* 1987] Z. H. Cho, O. Nalcioglu, and H. W. Park. Methods and algorithms for Fourier transform nuclear magnetic resonance tomography. *Journal of the Optical Society of America A*, 4(5):923–932, May 1987. Special Issue on Medical Imaging.

[Chommeloux *et al.* 1986] L. Chommeloux, C. Pichot, and J. C. Bolomey. Electromagnetic modeling for microwave imaging of cylindrical buried inhomogeneities. *IEEE Transactions on Microwave Theory and Techniques*, MTT-34(10):1064–1076, October 1986.

[Clough and Barrett 1983] A. V. Clough and H. H. Barrett. Attenuated Radon and Abel transforms. *Journal of the Optical Society of America*, 72(11):1590–1595, November 1983.

[Colton 1984] D. Colton. The inverse scattering problem for time-harmonic acoustic waves. *SIAM Review*, 26(3):323–350, July 1984.

[Colton and Monk 1987] D. Colton and P. Monk. The inverse scattering problem for time-harmonic acoustic waves in a penetrable medium. *Quarterly Journal of Mechanics and Applied Mathematics*, 40(2):184–212, May 1987.

[Connolly 1985] T. J. Connolly. *Inverse Problems for an Elliptic Equation.* Master's thesis, Department of Mathematics, University of Canterbury, Christchurch, New Zealand, 1985.

[Connolly and Wall 1987] T. J. Connolly and D. J. N. Wall. *An Inverse Problem, with Boundary Measurements for the Steady State Diffusion Equation.* Research Report 40, Department of Mathematics, University of Canterbury, Christchurch, New Zealand, October 1987.

[Connolly *et al.* 1985] T. J. Connolly, D. J. N. Wall, and R. H. T. Bates. Inverse problems and the Newton-Kantorovich method. In A. J. Devaney and R. H. T. Bates, editors, *Inverse Optics II*, pages 30–34, August 1985. Proceedings SPIE volume 558 (the Society of Photo-Optical Instrumentation Engineers).

[Copson 1975] E. T. Copson. *Partial Differential Equations.* Cambridge University Press, Cambridge, 1975.

[Cormack 1963] A. M. Cormack. Representation of a function by its line integrals, with some radiological applications. *Journal of Applied Physics*, 34(9):2722–2727, September 1963.

[Cormack 1964] A. M. Cormack. Representation of a function by its line integrals, with some radiological applications II. *Journal of Applied Physics*, 35(10):2908–2913, October 1964.

[Cornbleet 1983] S. Cornbleet. Geometrical optics reviewed: a new light on an old subject. *Proceedings IEEE*, 71(4):471–502, April 1983.

[Corones and Krueger 1983] J. P. Corones and R. J. Krueger. Higher order parabolic approximations to time-independent wave equations. *Journal of Mathematical Physics*, 24(9):2301–2304, September 1983.

[Craig and Brown 1986] I. J. D. Craig and J. C. Brown. *Inverse Problems in Astronomy.* Adam Hilger, Bristol BS1 6NX, England, 1986.

[Dainty 1974] J. C. Dainty. The transfer function, signal to noise ratio, and limiting magnitude in stellar speckle interferometry. *Monthly Notices of the Royal Astronomical Society*, 169(3):631–641, December 1974.

[Dainty 1975a] J. C. Dainty, editor. *Laser Speckle and Related Phenomena. Topics in Applied Physics volume 9*, Springer-Verlag, Berlin, 1975.

[Dainty 1975b] J. C. Dainty. Stellar speckle interferometry. In J. C. Dainty, editor, *Laser Speckle and Related Phenomena*, chapter 7, pages 255–280, Springer-Verlag, Berlin, 1975.

[Dainty and Fiddy 1984] J. C. Dainty and M. A. Fiddy. The essential role of prior information in phase retrieval. *Optica Acta*, 31(3):325–330, March 1984.

[Dallas and Wagner 1987] W. J. Dallas and R. F. Wagner, editors. Special issue on medical imaging. *Journal of the Optical Society of America A*, 4(5):891–992, May 1987.

[Datta and Bandyopadhyay 1985] A. N. Datta and B. Bandyopadhyay. An improved SIRT-style reconstruction algorithm for microwave tomography. *IEEE Transactions on Biomedical Engineering*, BME-32(9):719–722, September 1985.

[Davison 1983] M. E. Davison. The ill-conditioned nature of the limited angle tomography problem. *SIAM Journal of Applied Mathematics*, 43(2):428–448, April 1983.

[DeFacio and Rose 1985] B. DeFacio and J. H. Rose. Inverse-scattering theory for the non-spherically-symmetric three-dimensional plasma wave equation. *Physical Review A : General Physics*, 31(2):897–902, February 1985.

[Deighton *et al.* 1985] H. V. Deighton, M. S. Scivier, and M. A. Fiddy. Solution of the two-dimensional phase retrieval problem. *Optics Letters*, 10:250–251, 1985.

[Dennis and Schnabel 1983] J. E. Dennis and R. B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations. Prentice-Hall Series in Computational Mathematics*, Prentice-Hall, Engelwood Cliffs, New Jersey, 1983.

[Derenzo 1977] S. Derenzo. Positron ring cameras for emission-computed tomography. *IEEE Transactions on Nuclear Science*, NS-24(2):881–885, April 1977.

[Derenzo and Budinger 1977] S. E. Derenzo and T. F. Budinger. Resolution limit for positron-imaging devices. *Journal of Nuclear Medicine*, 18(5):491–492, 1977.

[Derenzo *et al.* 1977] S. E. Derenzo, T. F. Budinger, J. L. Cahoon, R. H. Huesman, and H. G. Jackson. High resolution computed tomography of positron emitters. *IEEE Transactions on Nuclear Science*, NS-24(1):544–558, February 1977.

[DeSanto 1977] J. A. DeSanto. Relation between the solutions of the Helmholtz and parabolic equations for sound propagation. *Journal of the Acoustical Society of America*, 62(2):295–297, August 1977.

[Deschamps and Cabayan 1972] G. A. Deschamps and H. S. Cabayan. Antenna synthesis and solution of inverse problems by regularisation methods. *IEEE Transactions on Antennas and Propagation*, AP-20(3):268–274, May 1972.

[Deuflhard and Hairer 1983] P. Deuflhard and E. Hairer, editors. *Numerical Treatment of Inverse Problems in Differential and Integral Equations: Proceedings of an International Workshop, Heidelberg, Federal Republic of Germany, August 30–September 3, 1982*, Birkhäuser Verlag, Boston, 1983.

[Devaney 1982] A. J. Devaney. A filtered backpropagation algorithm for diffraction tomography. *Ultrasonic Imaging*, 4(4):336–350, October 1982.

[Devaney 1983] A. J. Devaney. A computer simulation study of diffraction tomography. *IEEE Transactions on Biomedical Engineering*, BME-30(7):377–386, July 1983. Special Issue on Medical Ultrasound.

[Devaney 1985a] A. J. Devaney. Generalized projection-slice theorem for fan beam diffraction tomography. *Ultrasonic Imaging*, 7(3):264–275, July 1985.

[Devaney 1985b] A. J. Devaney, editor. Special issue on inverse problems in propagation and scattering. *Journal of the Optical Society of America A*, 2(11):1901–2061, November 1985.

[Devaney 1986] A. J. Devaney. Reconstructive tomography with diffracting wavefields. *Inverse Problems*, 2(2):161–183, May 1986.

[Devaney and Beylkin 1984] A. J. Devaney and G. Beylkin. Diffraction tomography using arbitrary transmitter and receiver surfaces. *Ultrasonic Imaging*, 6(2):181–193, April 1984.

[Devaney and Sherman 1982] A. J. Devaney and G. C. Sherman. Non-uniqueness in inverse source and scattering problems. *IEEE Transactions on Antennas and Propagation*, AP-30(5):1034–1037, September 1982.

[Devaney and Wolf 1973] A. J. Devaney and E. Wolf. Radiating and non-radiating classical current distributions and the fields they generate. *Physical Review D, Particles and Fields*, 8(4):1044–1047, 15 August 1973.

[Dines and Lytle 1979a] K. A. Dines and R. J. Lytle. Computerized geophysical tomography. *Proceedings IEEE*, 67(7):1065–1073, July 1979. See [Wait 1979,Dines and Lytle 1979b].

[Dines and Lytle 1979b] K. A. Dines and R. J. Lytle. Correction to 'Computerized geophysical tomography'. *Proceedings IEEE*, 67(12):1679, December 1979. See [Dines and Lytle 1979a].

[Dines and Lytle 1981] K. A. Dines and R. J. Lytle. Analysis of electrical conductivity imaging. *Geophysics*, 46(7):1025–1036, July 1981.

[Eftimiu and Huddleston 1984] C. Eftimiu and P. L. Huddleston. Reconstruction of a spherical scatterer from its natural frequencies. *IEEE Transactions on Antennas and Propagation*, AP-32(7):694–698, July 1984.

[Ehrenberg 1986] J. E. Ehrenberg, editor. Special issue on ocean acoustic remote sensing. *IEEE Journal of Oceanic Engineering*, OE-11(1):1–108, January 1986.

[Einziger and Felsen 1982] P. D. Einziger and L. B. Felsen. Evanescent waves and complex rays. *IEEE Transactions on Antennas and Propagation*, AP-30(4):594–605, July 1982.

[Erf 1978] R. K. Erf, editor. *Speckle Metrology*. Academic Press, New York, 1978.

[Fawcett and Keller 1985] J. Fawcett and H. B. Keller. Three-dimensional ray tracing and geophysical inversion in layered media. *SIAM Journal of Applied Mathematics*, 45(3):492–501, June 1985.

[Ferwerda 1978] H. A. Ferwerda. The phase reconstruction problem for wave amplitudes and coherence functions. In H. P. Baltes, editor, *Inverse Source Problems in Optics*, chapter 2, pages 13–39, Springer-Verlag, Berlin, 1978. See [Baltes 1978].

[Fienup 1978] J. R. Fienup. Reconstruction of an object from the modulus of its Fourier transform. *Optics Letters*, 3(1):27–29, July 1978.

[Fienup 1979] J. R. Fienup. Space object imaging through the turbulent atmosphere. *Optical Engineering*, 18(5):529–534, September-October 1979.

[Fienup 1980] J. R. Fienup. Iterative method applied to image reconstruction and to computer generated holograms. *Optical Engineering*, 19(3):297–305, May-June 1980.

[Fienup 1982] J. R. Fienup. Phase retrieval algorithms: A comparison. *Applied Optics*, 21(15):2758–2769, 1 August 1982.

[Fienup 1983a] J. R. Fienup, editor. Special issue on signal recovery. *Journal of the Optical Society of America*, 73(11):1412–1611, November 1983.

[Fienup 1983b] J. R. Fienup. Reconstruction of objects having latent reference points. *Journal of the Optical Society of America*, 73(11):1421–1426, November 1983. Special Issue on Signal Recovery.

[Fienup 1984] J. R. Fienup. Comparison of phase retrieval algorithms. In T. S. Huang, editor, *Advances in Computer Vision and Image Processing*, chapter 4, pages 191–225, JAI Press, Greenwich, Connecticut 06830, 1984.

[Fienup 1987] J. R. Fienup. Reconstruction of a complex-valued object from the modulus of its Fourier transform using a support constraint. *Journal of the Optical Society of America A*, 4(1):118–123, January 1987. Special Issue on Signal Recovery.

[Fienup and Rushforth 1987] J. R. Fienup and C. Rushforth, editors. Special issue on signal recovery. *Journal of the Optical Society of America A*, 4(1):101–304, January 1987.

[Fienup and Wackerman 1986] J. R. Fienup and C. C. Wackerman. Phase-retrieval stagnation problems and solutions. *Journal of the Optical Society of America A*, 3(11):1897–1907, November 1986.

[Fischer and Langenberg 1984] M. Fischer and K. J. Langenberg. Limitations and defects of certain inverse scattering theories. *IEEE Transactions on Antennas and Propagation*, AP-32(10):1080–1088, October 1984.

[Fishman and McCoy 1984a] L. Fishman and J. J. McCoy. Derivation and application of extended wave theories. I. The factored Helmholtz equation. *Journal of Mathematical Physics*, 25(2):285–296, February 1984.

[Fishman and McCoy 1984b] L. Fishman and J. J. McCoy. Derivation and application of extended wave theories. II. Path integral representations. *Journal of Mathematical Physics*, 25(2):297–308, February 1984.

[Fishman and McCoy 1984c] L. Fishman and J. J. McCoy. Factorization, path integral representations, and the construction of direct and inverse wave propagation theories. *IEEE Transactions on Geoscience and Remote Sensing*, GE-22(6):682–692, November 1984. Special Issue on the 1984 International Geoscience and Remote Sensing Symposium (IGARSS 1983): Remote Sensing: Extending Man's Horizon.

[Fletcher 1980] R. Fletcher. *Practical Methods of Optimization.* Volume 1: Unconstrained Optimization, Wiley, Chichester, England, 1980.

[Fletcher 1981] R. Fletcher. *Practical Methods of Optimization.* Volume 2: Constrained Optimization, Wiley, Chichester, England, 1981.

[Floyd *et al.* 1984] C. E. Floyd, R. J. Jaszczak, C. C. Harris, and R. E. Coleman. Energy and spatial distribution of multiple order Compton scatter in SPECT: A Monte Carlo investigation. *Physics in Medicine and Biology*, 29(10):1217–1230, October 1984.

[Floyd Jr. *et al.* 1985] C. E. Floyd, Jr., R. J. Jaszczak, K. L. Greer, and R. E. Coleman. Deconvolution of Compton scatter in SPECT. *Journal of Nuclear Medicine*, 26(4):403–408, April 1985.

[Fox 1950] C. Fox. *An Introduction to the Calculus of Variations.* Oxford University Press, London, 1950.

[Freeman *et al.* 1988] J. D. Freeman, J. C. Christou, D. L. McCarthy, and M. L. Cobb. A comparison of phase retrieval algorithms applied to infrared astronomical speckle data. In *Proceedings SPIE 828*, 1988. Paper number 7.

[Fright 1984] W. R. Fright. *The Fourier Phase Problem.* PhD thesis, Department of Electrical Engineering, University of Canterbury, Christchurch, New Zealand, 1984.

[Fritsch 1965] P. C. Fritsch, editor. Special issue on radar reflectivity. *Proceedings IEEE*, 53(8):770–1137, August 1965.

[Garabedian 1964] P. R. Garabedian. *Partial Differential Equations.* Wiley, New York, 1964.

[Garden 1984] K. L. Garden. *An Overview of Computed Tomography.* PhD thesis, Department of Electrical Engineering, University of Canterbury, Christchurch, New Zealand, 1984.

[Gardenier *et al.* 1986] P. H. Gardenier, C. A. Lim, D. G. H. Tan, and R. H. T. Bates. Aperture distribution phase from single radiation pattern measurement via Gerchberg-Saxton algorithm. *Electronics Letters*, 22(2):113–115, 16 January 1986.

[Garrett and Smithson 1987] J. A. Garrett and P. H. Smithson. Conventional X-ray imaging. *IEE Proceedings Part A*, 134(2):107–114, February 1987. Special Issue on Medical Imaging.

[Gerchberg 1986] R. W. Gerchberg. The lock problem in the Gerchberg-Saxton algorithm for phase retrieval. *Optik*, 74(3):91–93, October 1986.

[Gerchberg and Saxton 1971] R. W. Gerchberg and W. O. Saxton. Phase determination from image and diffraction plane pictures in the electron microscope. *Optik*, 34(3):275–284, December 1971.

[Gerchberg and Saxton 1972] R. W. Gerchberg and W. O. Saxton. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik*, 35(2):237–246, April 1972.

[Ghione *et al.* 1984] G. Ghione, I. Montrosset, and L. B. Felsen. Complex ray analysis of radiation from large apertures with tapered illumination. *IEEE Transactions on Antennas and Propagation*, AP-29(7):684–693, July 1984.

[Gilbert 1972] P. Gilbert. Iterative methods for the three-dimensional reconstruction of an object from projections. *Journal of Theoretical Biology*, 36(1):105–117, July 1972.

[Gill *et al.* 1981] P. E. Gill, W. Murray, and M. H. Wright. *Practical Optimization.* Academic Press, London, 1981.

[Gladwell 1984] G. M. L. Gladwell. The inverse problem for the vibrating beam. *Proceedings of the Royal Society of London A: Mathematical and Physical Sciences*, A393(1805):277–295, June 1984.

[Glover and Sharp 1977] G. H. Glover and J. C. Sharp. Reconstruction of ultrasound propagation speed distributions in soft tissue: Time-of-flight tomography. *IEEE Transactions on Sonics and Ultrasonics*, SU-24(4):229–234, July 1977.

[Goodman 1968] J. W. Goodman. *Introduction to Fourier Optics.* McGraw-Hill, San Francisco, 1968.

[Goodman 1975] J. W. Goodman. Statistical properties of laser speckle patterns. In J. C. Dainty, editor, *Laser Speckle and Related Phenomena*, chapter 2, pages 9–75, Springer-Verlag, Berlin, 1975.

[Gordon 1974] R. Gordon. A tutorial on ART (algebraic reconstruction techniques). *IEEE Transactions on Nuclear Science*, NS-21(3):78–93, June 1974. Special Issue on Physical and Computational Aspects of 3-Dimensional Image Reconstruction.

[Gordon *et al.* 1970] R. Gordon, R. Bender, and G. T. Herman. Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and x-ray photography. *Journal of Theoretical Biology*, 29(3):471–481, December 1970.

[Greenleaf 1983] J. F. Greenleaf. Computerized tomography with ultrasound. *Proceedings IEEE*, 71(3):330–337, March 1983. Special Issue on Computerized Tomography.

[Hämmerlin and Hoffmann 1983] G. Hämmerlin and K. H. Hoffmann, editors. *Improperly Posed Problems and Their Numerical Treatment: Conference Held at the Mathematisches Forschungsinstitut, Oberwolfach, September 26–October 2, 1982*, Birkhäuser Verlag, Basel, 1983.

[Harrington 1968] R. F. Harrington. *Field Computation by Moment Methods.* MacMillan, New York, 1968.

[Hashaby and Mittra 1987] T. M. Hashaby and R. Mittra. Review of some inverse methods in electromagnetics. *Journal of the Optical Society of America A*, 4(1):281–291, January 1987. Special Issue on Signal Recovery.

[Hayner and Jenkins 1984] D. A. Hayner and W. K. Jenkins. The missing cone problem in computer tomography. In T. S. Huang, editor, *Advances in Computer Vision and Image Processing volume 1*, chapter 2, pages 83–144, JAI Press, Greenwich, Connecticut, 1984. Image Reconstruction from Incomplete Observations.

[Heike 1986] U. Heike. Single-photon emission computed tomography by inverting the attenuated Radon transform with least squares collocation. *Inverse Problems*, 2(3):307–330, August 1986.

[Herman 1979] G. T. Herman, editor. *Image Reconstruction from Projections. Implementation and Applications. Topics in Applied Physics volume 32*, Springer-Verlag, Berlin, 1979.

[Herman 1983] G. T. Herman, editor. Special issue on computerized tomography. *Proceedings IEEE*, 71(3):291–435, March 1983.

[Herman and Lewitt 1979] G. T. Herman and R. M. Lewitt. Overview of image reconstruction from projections. In G. T. Herman, editor, *Image Reconstruction from Projections. Implementation and Applications*, chapter 1, pages 1–8, Springer-Verlag, Berlin, 1979. Topics in Applied Physics volume 32.

[Herstein 1975] I. N. Herstein. *Topics in Algebra.* Wiley, New York, second edition, 1975.

[Hinshaw and Lent 1983] W. S. Hinshaw and A. H. Lent. An introduction to NMR imaging: from the Bloch equation to the imaging equation. *Proceedings IEEE*, 71(3):338–350, March 1983. Special Issue on Computerized Tomography.

[Hoenders 1978] B. J. Hoenders. The uniqueness of inverse problems. In H. P. Baltes, editor, *Inverse Source Problems in Optics*, chapter 3, pages 41–82, Springer-Verlag, Berlin, 1978. See [Baltes 1978].

[Howard Jr. 1984] A. Q. Howard, Jr., editor. Special issue on electromagnetics in applied geophysics. *IEEE Transactions on Geoscience and Remote Sensing*, GE-22(1):2–87, January 1984.

[Hsieh and Wee 1976] R. C. Hsieh and W. G. Wee. On methods of three-dimensional reconstruction from a set of radioisotope scintigrams. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-6(12):854–862, December 1976.

[Ishimaru 1978] A. Ishimaru. *Wave Propagation and Scattering in Random Media.* Volume 2, Academic Press, New York, 1978.

[Jackson 1975] J. D. Jackson. *Classical Electrodynamics.* Wiley, New York, second edition, 1975.

[James 1976] G. L. James. *Geometrical Theory of Diffraction for Electromagnetic Waves.* Peter Peringrus, Stevenage, England, 1976. IEE electromagnetic wave series; 1.

[James 1986] B. D. James. *Vector Diffraction Tomography Using Radon Transform Techniques in Computer Assisted Electromagnetic Imaging.* Master's thesis, Department of Electrical Engineering and Computer Science, University of Illinois at Chicago, July 1986.

[James *et al.* 1985] B. D. James, C. W. Yang, and W. M. Boerner. Extension of scalar to vector diffraction propagation tomography — A computer numerical approach. In W. M. Boerner, editor, *Inverse Methods in Electromagnetic Imaging Part 1*, pages 1147–1164, Reidel, Dordrecht, Holland, 1985. Proceedings of the NATO Advanced Research Workshop, Bad Windsheim, Franconia, Federal Republic of Germany, 18–24 September 1983.

[Jaunzemis 1967] W. Jaunzemis. *Continuum Mechanics.* MacMillan, New York, 1967.

[John 1978] F. John. *Partial Differential Equations.* Springer-Verlag, New York, third edition, 1978.

[Johnson and Tracy 1983] S. A. Johnson and M. L. Tracy. Inverse scattering solutions by a sinc basis, multiple source, moment method — Part I: Theory. *Ultrasonic Imaging*, 5(4):361–375, October 1983.

[Johnson *et al.* 1975] S. A. Johnson, J. F. Greenleaf, W. A. Samayoa, F. A. Duck, and J. Sjostrand. Reconstruction of three-dimensional velocity fields and other parameters by acoustic ray tracing. In *IEEE Ultrasonics Symposium*, pages 46–51, 1975. Catalog number 75 CHO 994-4SU.

[Johnson *et al.* 1984] S. A. Johnson, Y. Zhou, M. K. Tracy, M. J. Berggren, and F. Stenger. Inverse scattering solutions by a sinc basis, multiple source, moment method — Part III: Fast algorithms. *Ultrasonic Imaging*, 6(1):103–116, January 1984. The third author's name appears in the journal as above but is probably meant to be M. L. Tracy [Johnson and Tracy 1983,Tracy and Johnson 1983].

[Jones 1964] D. S. Jones. *The Theory of Electromagnetism.* Pergamon Press, New York, 1964.

[Jones 1985] D. S. Jones. Note on a uniqueness theorem of Schiffer. *Applicable Analysis*, 19:181–188, 1985.

[Jordan and Ahn 1979] A. K. Jordan and S. Ahn. Inverse scattering theory and profile reconstruction. *Proceedings IEE*, 126(10):945–950, October 1979.

[Joseph and Schulz 1980] P. M. Joseph and R. A. Schulz. View sampling requirements in fan beam computed tomography. *Medical Physics*, 7(6):692–702, November/December 1980.

[Joseph and Spital 1978] P. M. Joseph and R. D. Spital. A method for correcting bone induced artifacts in computed tomography scanners. *Journal of Computer Assisted Tomography*, 2:100–108, January 1978.

[Joseph and Spital 1982] P. M. Joseph and R. D. Spital. The effects of scatter in x-ray computed tomography. *Medical Physics*, 9(4):464–472, July/August 1982.

[Kac 1966] M. Kac. Can one hear the shape of a drum? *American Mathematical Monthly*, 73(4 Part II):1–23, April 1966.

[Kak 1979] A. C. Kak. Computerized tomography with X-ray, emission, and ultrasound sources. *Proceedings IEEE*, 67(9):1245–1272, September 1979. Special Issue on Technology and Health Care.

[Kaveh *et al.* 1982] M. Kaveh, M. Soumekh, Z. Q. Lu, R. K. Mueller, and J. F. Greenleaf. Further results on diffraction tomography using Rytov's approximation. In E. A. Ash and C. R. Hill, editors, *Acoustical Imaging 12—Proceedings of the Twelfth International Symposium on Acoustical Imaging*, pages 599–608, Plenum Press, New York, London, 1982.

[Keller 1962] J. B. Keller. Geometrical theory of diffraction. *Journal of the Optical Society of America*, 52(2):116–130, 1962.

[Keller 1969] J. B. Keller. Accuracy and validity of the Born and Rytov approximations. *Journal of the Optical Society of America*, 59(8):1003–1004, August 1969.

[Keyes 1987] W. I. Keyes. Radionuclide imaging. *IEE Proceedings Part A*, 134(2):161–170, February 1987. Special Issue on Medical Imaging.

[Kim and Wolf 1986] K. Kim and E. Wolf. Non-radiating monochromatic sources and their fields. *Optics Communications*, 59(1):1–6, 1 August 1986.

[Kline and Kay 1965] M. Kline and I. W. Kay. *Electromagnetic Theory and Geometrical Optics.* Wiley-Interscience, New York, 1965. Pure and Applied Mathematics, a Series of Monographs and Texts, volume 12.

[Knoll 1983] G. F. Knoll. Single-photon emission computed tomography. *Proceedings IEEE*, 71(3):320–329, March 1983. Special Issue on Computerized Tomography.

[Kock 1973] W. E. Kock. *Radar, Sonar, and Holography, an Introduction.* Academic Press, New York, 1973.

[Kohn and Vogelius 1984] R. V. Kohn and M. Vogelius. Determining conductivity by boundary measurements. *Communications on Pure and Applied Mathematics*, 37:289–298, 1984.

[Kohn and Vogelius 1985] R. V. Kohn and M. Vogelius. Determining conductivity by boundary measurements II. Interior results. *Communications on Pure and Applied Mathematics*, 38:643–667, 1985.

[Kristensson and Krueger 1987] G. Kristensson and R. J. Krueger. Direct and inverse scattering in the time domain for a dissipative wave equation. III. Scattering operators in the presence of a phase velocity mismatch. *Journal of Mathematical Physics*, 28(2):360–370, February 1987.

[Kristensson and Vogel 1986] G. Kristensson and C. R. Vogel. Inverse problems for acoustic waves using the penalised likelihood method. *Inverse Problems*, 2(4):461–479, November 1986.

[Kuc 1984] R. Kuc. Estimating acoustic attenuation from reflected ultrasound signals: Comparison of spectral-shift and spectral-difference approaches. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-32(1):1–6, February 1984.

[Labeyrie 1970] A. Labeyrie. Attainment of diffraction limited resolution in large telescopes by fourier analysing speckle patterns in star images. *Astronomy and Astrophysics*, 6(1):85–87, May 1970.

[LaHaie 1986] I. J. LaHaie. Uniqueness of the inverse source problem for quasi-homogeneous, partially coherent sources. *Journal of the Optical Society of America A*, 3(7):1073–1079, July 1986.

[Lalor and Wolf 1972] E. Lalor and E. Wolf. Exact solutions of the equations of molecular optics for refraction and reflection of an electromagnetic plane wave on a semi-infinite dielectric. *Journal of the Optical Society of America*, 62(10):1165–1174, October 1972.

[Lan *et al.* 1985] C. Q. Lan, K. K. Zu, and G. Wade. Limited angle diffraction tomography and its application to planar scanning systems. *IEEE Transactions on Sonics and Ultrasonics*, SU-32(1):9–16, January 1985.

[Lane 1987] R. G. Lane. Recovery of complex images from Fourier magnitude. *Optics Communications*, 63(1):6–10, 1 July 1987.

[Lane and Bates 1987a] R. G. Lane and R. H. T. Bates. Automatic multi-dimensional deconvolution. *Journal of the Optical Society of America A*, 4(1):180–188, January 1987. Special Issue on Signal Recovery.

[Lane and Bates 1987b] R. G. Lane and R. H. T. Bates. Relevance for blind deconvolution of recovering Fourier magnitude from phase. *Optics Communications*, 63(1):11–14, 1 July 1987.

[Lane *et al.* 1987] R. G. Lane, W. R. Fright, and R. H. T. Bates. Direct phase retrieval. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-35(4):520–526, April 1987.

[Larsson and Israelsson 1982] S. Larsson and A. Israelsson. Considerations on system design, implementation and computer processing in SPECT. *IEEE Transactions on Nuclear Science*, NS-29(4):1331–1342, August 1982.

[Lawrence 1985] M. H. Lawrence. Ray theory modeling applied to low-frequency acoustic interaction with horizontally stratified ocean bottoms. *Journal of the Acoustical Society of America*, 78(2):649–658, August 1985.

[Lax and Phillips 1967] P. D. Lax and R. S. Phillips. *Scattering Theory. Pure and Applied Mathematics, a series of monographs and textbooks; volume 26*, Academic Press, New York, 1967.

[Lee 1984] H. Lee. Formulation of the generalized backward-projection method for acoustical imaging. *IEEE Transactions on Sonics and Ultrasonics*, SU-31(3):157–161, May 1984.

[Lee and Marcuvitz 1986] S. Y. Lee and N. Marcuvitz. Quasiparticle description of wave propagation and reflection in inhomogeneous media. *IEEE Transactions on Antennas and Propagation*, AP-34(5):613–615, May 1986.

[Leeman *et al.* 1985] S. Leeman, M. A. Fiddy, and L. Zapalowski. Born versus Rytov: is the debate over? In A. J. Devaney and R. H. T. Bates, editors, *Inverse Optics II*, pages 11–14, August 1985. Proceedings SPIE volume 558 (the Society of Photo-Optical Instrumentation Engineers).

[Lewin 1970] L. Lewin. On the restricted validity of point-matching techniques. *IEEE Transactions on Microwave Theory and Techniques*, MTT-18(12):1041–1047, December 1970.

[Lewis 1969] R. M. Lewis. Physical optics inverse diffraction. *IEEE Transactions on Antennas and Propagation*, AP-17(3):308–314, May 1969.

[Lewis *et al.* 1982] M. H. Lewis, J. T. Willerson, S. E. Lewis, F. J. Bonte, and R. W. Parkey. Attenuation compensation in single-photon emission tomography: A comparative evaluation. *Journal of Nuclear Medicine*, 23(12):1121–1127, 1982.

[Lewitt 1983] R. M. Lewitt. Reconstruction algorithms: Transform methods. *Proceedings IEEE*, 71(3):390–408, March 1983. Special Issue on Computerized Tomography.

[Lewitt and Bates 1978a] R. M. Lewitt and R. H. T. Bates. Image reconstruction from projections: I: General theoretical considerations. *Optik*, 50(1):19–33, February 1978.

[Lewitt and Bates 1978b] R. M. Lewitt and R. H. T. Bates. Image reconstruction from projections: IV: Projection completion methods (computational examples). *Optik*, 50(4):269–278, May 1978.

[Lewitt *et al.* 1978] R. M. Lewitt, R. H. T. Bates, and T. M. Peters. Image reconstruction from projections: II: Modified back-projection methods. *Optik*, 50(2):85–109, March 1978.

[Liu and Chen 1984] J. Q. Liu and Y. M. Chen. An iterative algorithm for solving inverse problems of two-dimensional diffusion equations. *SIAM Journal on Scientific and Statistical Computing*, 5(2):255–269, June 1984.

[Lytle and Dines 1980] R. J. Lytle and K. A. Dines. Iterative ray-tracing between boreholes for underground image reconstruction. *IEEE Transactions on Geoscience and Remote Sensing*, GE-18(3):234–240, July 1980.

[Lytle and Lager 1976] R. J. Lytle and D. L. Lager. Using the natural-frequency concept in remote probing of the earth. *Radio Science*, 11(3):199–209, March 1976.

[Malacara 1985] D. Malacara. Optical and electronic processing of medical images. In E. Wolf, editor, *Progress in Optics volume XXII*, chapter 1, pages 1–76, North-Holland Physics, Amsterdam, 1985.

[Marcuvitz 1980] N. Marcuvitz. Quasiparticle view of propagation. *Proceedings IEEE*, 68(11):1380–1395, November 1980.

[McCullough and Payne 1977] E. McCullough and J. T. Payne. X-ray-transmission computed tomography. *Medical Physics*, 4(2):85–98, March/April 1977.

[McKinnon 1980] G. C. McKinnon. *Contributions to Imaging.* PhD thesis, Department of Electrical Engineering, University of Canterbury, Christchurch, New Zealand, 1980.

[McKinnon and Bates 1980] G. C. McKinnon and R. H. T. Bates. A limitation on ultrasonic transmission tomography. *Ultrasonic Imaging*, 2(1):48–54, January 1980.

[McLaughlin 1984a] D. W. McLaughlin, editor. *Proceedings of the Symposium in Applied Mathematics of the American Mathematical Society and the Society for Industrial and Applied Mathematics, New York City, April 12–13, 1983*, American Mathematical Society, Rhode Island, 1984.

[McLaughlin 1984b] J. R. McLaughlin. On constructing solutions to an inverse Euler-Bernoulli beam problem. In F. Santosa, Y. H. Pao, W. W. Symes, and C. Holland, editors, *Inverse Problems of Acoustic and Elastic Waves*, pages 341–345, Society for Industrial and Applied Mathematics, Philadelphia, 1984.

[McLaughlin 1986] J. R. McLaughlin. Analytical methods for recovering coefficients in differential equations from spectral data. *SIAM Review*, 28(1):53–72, March 1986.

[McLaughlin and Rundell 1987] J. R. McLaughlin and W. Rundell. A uniqueness theorem for an inverse Sturm-Liouville problem. *Journal of Mathematical Physics*, 28(7):1471–1472, July 1987.

[Melton Jr. and Magnin 1984] H. E. Melton, Jr. and P. A. Magnin. A-mode speckle reduction with compound frequencies and compound bandwidths. *Ultrasonic Imaging*, 6(2):159–173, April 1984.

[Minard *et al.* 1985] R. A. Minard, B. S. Robinson, and R. H. T. Bates. Full-wave computed tomography. Part 3: Coherent shift-and-add imaging. *IEE Proceedings Part A*, 132(1):50–58, January 1985.

[Misell 1973] D. L. Misell. An examination of an iterative method for the solution of the phase problem in optics and electron optics: I. Test calculations. *Journal of Physics D: Applied Physics*, 6(18):2200–2216, 5 December 1973.

[Mittra 1985] R. Mittra, editor. Special issue on direct problems in propagation and scattering. *Journal of the Optical Society of America A*, 2(6):883–1006, June 1985.

[Moffat *et al.* 1981] D. L. Moffat, J. D. Young, A. A. Ksienski, H. C. Lim, and C. M. Rhoads. Transient response characteristics in identification and imaging. *IEEE Transactions on Antennas and Propagation*, AP-29(2):192–205, 1981. Special Issue on Inverse Methods in Electromagnetics.

[Morris 1985] D. Morris. Phase retrieval in the radio holography of reflector antennas and radio telescopes. *IEEE Transactions on Antennas and Propagation*, AP-33(7):749–755, July 1985.

[Morris 1986] P. G. Morris. *Nuclear Magnetic Resonance Imaging in Medicine and Biology*. Clarendon Press, Oxford, 1986.

[Morse and Feshbach 1953] P. M. Morse and H. Feshbach. *Methods of Theoretical Physics*. McGraw-Hill, New York, 1953. Volume 1 : Chapters 1 to 8; volume 2 : Chapters 9 to 13.

[Morse and Ingard 1968] P. M. Morse and K. U. Ingard. *Theoretical Acoustics*. McGraw-Hill, New York, 1968.

[Moses and Prosser 1986] H. E. Moses and R. T. Prosser. Initial conditions, sources, and currents for prescribed time-dependent acoustic and electromagnetic fields in three dimensions, part I: the inverse initial value problem. Acoustic and electromagnetic "bullets," expanding waves, and imploding waves. *IEEE Transactions on Antennas and Propagation*, AP-34(2):188–196, February 1986.

[Muehllehner and Colsher 1981] G. Muehllehner and J. Colsher. Single photon imaging: New instrumentation and techniques. In *Medical Radionuclide Imaging volume 1*, pages 173–198, International Atomic Energy Agency, Vienna, 1981. IAEA-SM-247/202.

[Mueller 1980] R. K. Mueller. Diffraction tomography I: the wave equation. *Ultrasonic Imaging*, 2(3):213–222, July 1980.

[Müller 1969] C. Müller. *Foundations of the Mathematical Theory of Electromagnetic Waves.* Springer-Verlag, New York, 1969.

[Nashed 1981] M. Z. Nashed. Operator-theoretic and computational approaches to ill-posed problems with applications to antenna theory. *IEEE Transactions on Antennas and Propagation*, AP-29(2):220–231, March 1981. Special Issue on Inverse Methods in Electromagnetics.

[Natterer 1983] F. Natterer. On the order of regularization methods. In G. Hämmerlin and K. H. Hoffmann, editors, *Improperly Posed Problems and Their Numerical Treatment: Conference Held at the Mathematisches Forschungsinstitut, Oberwolfach, September 26–October 2, 1982*, pages 189–203, Birkhäuser Verlag, Basel, 1983.

[Natterer 1985] F. Natterer. Fourier reconstruction in tomography. *Numerische Mathematik*, 47(3):343–353, November 1985.

[Newton 1966] R. G. Newton. *Scattering Theory of Waves and Particles.* McGraw-Hill, New York, 1966.

[Newton 1980] R. G. Newton. Inverse scattering. II. Three dimensions. *Journal of Mathematical Physics*, 21(7):1698–1715, July 1980. See [Newton 1981a].

[Newton 1981a] R. G. Newton. Erratum to 'Inverse scattering. II. Three dimensions'. *Journal of Mathematical Physics*, 22(3):631, March 1981. See [Newton 1980].

[Newton 1981b] R. G. Newton. Inversion of reflection data for layered media: A review of exact methods. *Geophysical Journal of the Royal Astronomical Society*, 65(1):191–215, April 1981.

[Newton 1981c] R. G. Newton. Inverse scattering. III. Three dimensions, continued. *Journal of Mathematical Physics*, 22(10):2191–2200, October 1981. See [Newton 1982b].

[Newton 1982a] R. G. Newton. Inverse scattering. IV. Three dimensions: generalized Marchenko construction with bound states, and generalized Gel'fand Levitan equations. *Journal of Mathematical Physics*, 23(4):594–604, April 1982.

[Newton 1982b] R. G. Newton. Erratum to 'Inverse scattering. III. Three dimensions, continued'. *Journal of Mathematical Physics*, 23(5):693, May 1982. See [Newton 1981c].

[Newton 1985] R. G. Newton. Relation between the Schrödinger equation and the plasma wave equation. *Physical Review A, General Physics*, 31(5):3305–3308, May 1985.

[Nievergelt 1986] Y. Nievergelt. Elementary inversion of Radon's transform. *SIAM Review*, 28(1):79–84, March 1986.

[Norton 1983] S. J. Norton. Generation of separate density and compressibility images in tissue. *Ultrasonic Imaging*, 5(3):240–252, July 1983.

[Norton 1985] S. J. Norton. Iterative reconstruction algorithms: Convergence as a function of spatial frequency. *Journal of the Optical Society of America A*, 2(1):6–13, January 1985.

[Norton and Linzer 1982] S. J. Norton and M. Linzer. Correcting for ray refraction in velocity and attenuation tomography: a perturbation approach. *Ultrasonic Imaging*, 3(2):201–233, July 1982.

[Ohyama *et al.* 1987] N. Ohyama, S. Ohki, S. Inoue, J. Tsujiuchi, and T. Honda. Discrete Radon transform in a continuous space. *Journal of the Optical Society of America A*, 4(2):318–324, February 1987.

[Oldenburg 1984] D. W. Oldenburg. An introduction to linear inverse theory. *IEEE Transactions on Geoscience and Remote Sensing*, GE-22(6):665–674, November 1984. Special Issue on the 1984 International Geoscience and Remote Sensing Symposium (IGARSS 1983): Remote Sensing: Extending Man's Horizon.

[Olver 1974] F. W. J. Olver. *Introduction to Asymptotics and Special Functions.* Academic Press, New York, 1974.

[O'Sullivan 1985] J. D. O'Sullivan. A fast sinc function gridding algorithm for Fourier inversion in computer tomography. *IEEE Transactions on Medical Imaging*, MI-4(4):200-207, December 1985.

[Pan and Kak 1983] S. X. Pan and A. C. Kak. A computational study of reconstruction algorithms for diffraction tomography. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-31(5):1262–1275, October 1983.

[Pao and Varatharajulu 1976] Y. H. Pao and V. Varatharajulu. Huygen's principle, radiation conditions, and formulas for the scattering of elastic waves. *Journal of the Acoustical Society of America*, 59(6):1361–1371, June 1976.

[Paoloni 1986] F. J. Paoloni. The effects of attenuation on the Born reconstruction procedure for microwave diffraction tomography. *IEEE Transactions on Microwave Theory and Techniques*, MTT-34(3):366–368, March 1986.

[Paoloni 1987] F. J. Paoloni. Implementation of microwave diffraction tomography for measurement of dielectric constant distribution. *IEE Proceedings Part H*, H134(1):25–29, February 1987.

[Pattanayak and Wolf 1972] D. N. Pattanayak and E. Wolf. General form and a new interpretation of the Ewald-Oseen extinction theorem. *Optics Communications*, 6(3):217–220, November 1972.

[Pauling 1985] L. Pauling. Apparent icosahedral symmetry is due to directed twinning of cubic crystals. *Nature*, 317(6037):512–514, 10–16 October 1985.

[Peng and Stark 1987] H. Peng and H. Stark. Direct Fourier reconstruction in fan-beam tomography. *IEEE Transactions on Medical Imaging*, MI-6(3):209–219, September 1987.

[Peterson and Ström 1975] B. Peterson and S. Ström. Matrix formulation of acoustic scattering from multilayered scatterers. *Journal of the Acoustical Society of America*, 57(1):2–13, January 1975.

[Porter 1970] R. P. Porter. Diffraction-limited, scalar image formation with holograms of arbitrary shape. *Journal of the Optical Society of America*, 60(8):1051–1059, August 1970.

[Porter and Devaney 1982a] R. P. Porter and A. J. Devaney. Holography and the inverse source problem. *Journal of the Optical Society of America*, 72(3):327–330, March 1982.

[Porter and Devaney 1982b] R. P. Porter and A. J. Devaney. Generalised holography and computational solution to inverse scattering problems. *Journal of the Optical Society of America*, 72(12):1707–1713, December 1982.

[Price 1979] L. R. Price. Electrical impedance computed tomography (ICT): A new CT imaging technique. *IEEE Transactions on Nuclear Science*, NS-26(2 Part 2):2736–2739, April 1979.

[Prosser 1972] R. T. Prosser. The inverse problem for radiation scattering. In L. Colin, editor, *Mathematics of Profile Inversion*, pages 6–18–6–27, August 1972. NASA Technical Memorandum X-62,150. Proceedings of a Workshop held at Ames Research Center, Moffett Field, California 94035, July 12–16, 1971.

[Radon 1917] J. Radon. On the determination of functions from their integral values along certain manifolds. *Berichte der Sachsischen Akademie der Wissenschatt*, 69:262–277, April 1917. Translation from German by P. C. Parks in *IEEE Transactions on Medical Imaging*, MI-5(4):170–176, December 1986.

[Rall 1969] L. B. Rall. *Computational Solution of Nonlinear Operator Equations.* Wiley, New York, 1969.

[Ramachandran and Srinivasan 1970] G. N. Ramachandran and R. Srinivasan. *Fourier Methods in Crystallography.* Wiley-Interscience, New York, 1970.

[Ramm 1986] A. G. Ramm. Inverse scattering for geophysical problems when the background is variable. *Journal of Mathematical Physics*, 27(11):2687–2689, November 1986.

[Ramm 1987] A. G. Ramm. Necessary and sufficient conditions on the scattering data for the potential to be in l^2 for the Schrödinger operator on the half-line. *Inverse Problems*, 3(4):L71–L75, November 1987.

[Rea 1984] D. G. Rea, editor. Special issue on the 1984 International Geoscience and Remote Sensing Symposium (IGARSS 1983): Remote Sensing: Extending Man's Horizon. *IEEE Transactions on Geoscience and Remote Sensing*, GE-22(6):469–727, November 1984.

[Rivlin 1969] T. J. Rivlin. *An Introduction to the Approximation of Functions.* Blaisdell, Waltham, Massachusetts, 1969. A Blaisdell Book in Numerical Analysis and Computer Science.

[Robinson 1982] B. S. Robinson. *Speckle Processing for Ultrasonic Imaging.* PhD thesis, Department of Electrical Engineering, University of Canterbury, Christchurch, New Zealand, 1982.

[Robinson 1984] E. A. Robinson. Image reconstruction in exploration geophysics. *IEEE Transactions on Sonics and Ultrasonics*, SU-31(4):259–270, July 1984. Special Issue on Digital Acoustical Imaging.

[Robinson and Bates 1980] B. S. Robinson and R. H. T. Bates. Wideband ultrasonic diffraction measurements. *Australasian Physical and Engineering Sciences in Medicine*, 3(6):233–238, 1980.

[Robinson and Greenleaf 1986] B. S. Robinson and J. F. Greenleaf. The scattering of ultrasound by cylinders: implications for diffraction tomography. *Journal of the Acoustical Society of America*, 80(1):40–49, July 1986.

[Robinson and Knight 1981] D. E. Robinson and P. C. Knight. Computer reconstruction techniques in compound scan pulse-echo imaging. *Ultrasonic Imaging*, 3(3):217–234, July 1981.

[Roger 1981] A. Roger. Newton-Kantorovitch algorithm applied to an electromagnetic inverse problem. *IEEE Transactions on Antennas and Propagation*, AP-29(2):232–238, March 1981.

[Rose *et al.* 1984] J. H. Rose, M. Cheney, and B. DeFacio. The connection between time- and frequency-domain three-dimensional inverse scattering techniques. *Journal of Mathematical Physics*, 25(10):2995–3000, October 1984.

[Rose *et al.* 1985] J. H. Rose, M. Cheney, and B. DeFacio. Three-dimensional inverse scattering: plasma and variable velocity wave equations. *Journal of Mathematical Physics*, 26(11):2803–2813, November 1985.

[Rose *et al.* 1986] J. H. Rose, M. Cheney, and B. DeFacio. Determination of the wave field from scattering data. *Physical Review Letters*, 57(7):783–786, 18 August 1986.

[Rotenberg and Johns 1965] A. D. Rotenberg and H. E. Johns. Collimator efficiency and design. *Physics in Medicine and Biology*, 10(1):51–65, 1965.

[Rowland 1979] S. W. Rowland. Computer implementation of image reconstruction formulas. In G. T. Herman, editor, *Image Reconstruction from Projections. Implementation and Applications*, chapter 2, pages 9–79, Springer-Verlag, Berlin, 1979. Topics in Applied Physics volume 32.

[Sabatier 1978] P. C. Sabatier, editor. *Applied Inverse Problems. Lecture Notes in Physics volume 85*, Springer-Verlag, Berlin, 1978. Lectures presented at the RCP 264 "Etude Interdisciplinaire des Problèmes Inverses" sponsored by the Centre National de la Recherche Scientifique.

[Sabatier 1983] P. C. Sabatier. Theoretical considerations for inverse scattering. *Radio Science*, 18(1):1–18, January-February 1983.

[Santosa *et al.* 1984] F. Santosa, Y. H. Pao, W. W. Symes, and C. Holland, editors. *Proceedings of International Conference on Inverse Problems of Acoustic and Elastic Waves, Cornell University*, Philadelphia, 1984.

[Saxton 1978] W. O. Saxton. *Computer Techniques for Image Processing in Electron Microscopy.* Academic Press, New York, 1978. Advances in Electronics and Electron Physics *Supplement 10.*

[Schomberg 1978] H. Schomberg. An improved approach to reconstructive ultrasound tomography. *Journal of Physics D: Applied Physics*, 11:L181–L185, 1978.

[Seagar 1983] A. D. Seagar. *Probing with Low Frequency Electric Currents.* PhD thesis, Department of Electrical Engineering, University of Canterbury, Christchurch, New Zealand, 1983.

[Seagar and Bates 1985] A. D. Seagar and R. H. T. Bates. Full-wave computed tomography IV. Low-frequency electric current CT. *IEE Proceedings Part A*, A132(7):455–466, November 1985.

[Senior 1960] T. B. A. Senior. Impedance boundary conditions for imperfectly conducting surfaces. *Applied Scientific Research*, B8(5–6):413–436, 1960.

[Senior 1965] T. B. A. Senior. A survey of analytical techniques for cross-section estimation. *Proceedings IEEE*, 53(8):822–833, August 1965. Special Issue on Radar Reflectivity.

[Sezan and Stark 1984] M. I. Sezan and H. Stark. Tomographic image reconstruction from incomplete data by convex projections and direct Fourier inversion. *IEEE Transactions on Medical Imaging*, MI-3(2):91–98, June 1984.

[Shechtman *et al.* 1984] D. Shechtman, I. Blech, D. Gratias, and J. W. Cahn. Metallic phase with long-range orientational order and no translational symmetry. *Physical Review Letters*, 53(20):1951–1953, 12 November 1984.

[Sherman 1969] G. C. Sherman. Diffracted wave fields expressible by plane-wave expansions containing only homogeneous waves. *Journal of the Optical Society of America*, 59(6):697–711, June 1969.

[Sherman 1979] H. Sherman, editor. Special issue on technology and health care. *Proceedings IEEE*, 67(9):1188–1358, September 1979.

[Shilov 1968] G. E. Shilov. *Generalized Functions and Partial Differential Equations.* Gordon and Breach, New York, authorized English edition, 1968.

[Sieber 1983] A. J. Sieber, editor. Special issue on the 1983 International Geoscience and Remote Sensing Symposium (IGARSS 1982): Remote Sensing: The promise of Remote Sensing. *IEEE Transactions on Geoscience and Remote Sensing*, GE-21(3):228–405, July 1983.

[Siegmann *et al.* 1985] W. L. Siegmann, G. A. Kriegsmann, and D. Lee. A wide-angle three-dimensional parabolic wave equation. *Journal of the Acoustical Society of America*, 78(2):659–664, August 1985.

[Skolnik 1962] M. I. Skolnik. *Introduction to Radar Systems.* McGraw-Hill Kogakusha, Tokyo, international student edition, 1962.

[Slaney *et al.* 1984] M. Slaney, A. C. Kak, and L. E. Larsen. Limitations of imaging with first-order diffraction tomography. *IEEE Transactions on Microwave Theory and Techniques*, MTT-32(8):860–874, August 1984.

[Smith 1978] G. D. Smith. *Numerical Solutions of Partial Differential Equations. Oxford Applied Mathematics and Computing Science Series*, Clarendon Press, Oxford, second edition, 1978.

[Smith and Vozoff 1984] N. C. Smith and K. Vozoff. Two-dimensional DC resistivity inversion for dipole-dipole data. *IEEE Transactions on Geoscience and Remote Sensing*, GE-22(1):21–28, January 1984. Special Issue on Electromagnetics in Applied Geophysics.

[Soumekh 1986a] M. Soumekh. An improvement to the Rytov approximation in diffraction tomography. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, UFFC-33(4):394–401, July 1986.

[Soumekh 1986b] M. Soumekh. Image reconstruction techniques in tomographic imaging systems. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-34(4):952–962, August 1986.

[Soumekh and Kaveh 1986] M. Soumekh and M. Kaveh. A theoretical study of model approximation errors in diffraction tomography. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, UFFC-33(1):10–20, January 1986.

[Spencer 1980] A. J. M. Spencer. *Continuum Mechanics.* Longman, London, 1980.

[Spiesberger 1985] J. L. Spiesberger. Ocean acoustic tomography. *Journal of the Acoustical Society of America*, 77(1):83–100, January 1985.

[Stakgold 1979] I. Stakgold. *Green's Functions and Boundary Value Problems.* Wiley-Interscience, New York, 1979.

[Stanton 1969] L. Stanton. *Basic Medical Radiation Physics.* Appleton-Century-Crofts, New York, 1969.

[Stark *et al.* 1981] H. Stark, J. W. Woods, I. Paul, and R. Hingorani. An investigation of computerized tomography by direct Fourier inversion and optimum interpolation. *IEEE Transactions on Biomedical Engineering*, BME-28(7):496–505, July 1981.

[Straughan 1986] K. Straughan. NMR imaging: principles and current status. In R. P. Clark and M. R. Goff, editors, *Recent Developments in Medical and Physiological Imaging*, pages 49–59, Taylor and Francis, London, 1986. A supplement to *Journal of Medical Engineering & Technology.*

[Tan and Bates 1988] D. G. H. Tan and R. H. T. Bates. Generalised volume source formulation of diffraction and a sequence of approximations. To appear in *Wave Motion*, 1988.

[Tan *et al.* 1986] D. G. H. Tan, J. X. Qu, and R. H. T. Bates. Image reconstruction from projections: VIII: Effects of finite resolution and sampling of individual projections. *Optik*, 73(1):25–29, April 1986.

[Tan *et al.* 1987] D. G. H. Tan, J. X. Qu, K. L. Garden, and R. H. T. Bates. Allowing for variable resolution and constant attenuation in SPECT. *IEE Proceedings Part A*, A134(2):136–142, February 1987.

[Tanaka *et al.* 1984] E. Tanaka, H. Toyama, and H. Murayama. Convolutional image reconstruction for quantitative single photon emission computed tomography. *Physics in Medicine and Biology*, 29(12):1489–1500, December 1984.

[Thompson *et al.* 1979] C. J. Thompson, Y. L. Yamamoto, and E. Meyer. Positome II: A high efficiency positron imaging device for dynamic brain studies. *IEEE Transactions on Nuclear Science*, NS-26(1):583–589, February 1979.

[Thompson *et al.* 1986] C. J. Thompson, A. Dagher, E. Meyer, and A. C. Evans. Imaging performance of a dynamic positron imaging tomograph: Positome IIIp. *IEEE Transactions on Medical Imaging*, MI-5(4):183–198, December 1986.

[Thomson and Wood 1987] D. J. Thomson and D. H. Wood. A postprocessing method for removing phase errors in the parabolic approximation. *Journal of the Acoustical Society of America*, 82(1):224–232, July 1987.

[Tihonov 1963a] A. N. Tihonov. Solution of incorrectly formulated problems and the regularization method. *Soviet Mathematics (American Mathematical Society translation of Doklady, Academy of Sciences of the USSR)*, 4(4):1035–1038, July-August 1963.

[Tihonov 1963b] A. N. Tihonov. Regularization of incorrectly posed problems. *Soviet Mathematics (American Mathematical Society translation of Doklady, Academy of Sciences of the USSR)*, 4(6):1624–1627, November-December 1963.

[Tracy and Johnson 1983] M. L. Tracy and S. A. Johnson. Inverse scattering solutions by a sinc basis, multiple source, moment method — Part II: Numerical evaluation. *Ultrasonic Imaging*, 5(4):376–392, October 1983.

[Tretiak and Metz 1980] O. Tretiak and C. Metz. The exponential Radon transform. *SIAM Journal of Applied Mathematics*, 39(2):341–354, October 1980.

[Truesdell 1977] C. A. Truesdell. *A First Course in Rational Continuum Mechanics, volume 1: General Concepts.* Academic Press, New York, 1977. Pure and Applied Mathematics: A Series of Monographs and Textbooks volume 71.1.

[Ufimtsev 1971] Y. Ufimtsev. *Method of Edge Waves in the Physical Theory of Diffraction.* Sovyetskoye radio, Moscow, 1971. Translated by U.S. Air Force Foreign Technology Division, Wright-Patterson AFB, Ohio, USA.

[Varadan and Varadan 1980] V. K. Varadan and V. V. Varadan, editors. *Acoustic, Electromagnetic and Elastic Wave Scattering — Focus on the T-matrix approach.* Pergamon Press, New York, 1980.

[Varadan *et al.* 1978] V. K. Varadan, V. V. Varadan, and Y. H. Pao. Multiple scattering of elastic waves by cylinders of arbitrary cross section. *Journal of the Acoustical Society of America*, 63(5):1310–1319, May 1978.

[Varatharajulu and Pao 1976] V. Varatharajulu and Y. H. Pao. Scattering matrix for elastic waves. I. Theory. *Journal of the Acoustical Society of America*, 60(3):556–566, September 1976.

[Waag 1983] R. C. Waag, editor. Special issue on medical ultrasound. *IEEE Transactions on Biomedical Engineering*, BME-30(6):429–492, August 1983.

[Wade *et al.* 1984] G. Wade, H. Lee, and C. Schueler, editors. Special issue on digital acoustical imaging. *IEEE Transactions on Sonics and Ultrasonics*, SU-31(4):193–440, July 1984.

[Wait 1979] J. R. Wait, editor. Special issue on applications of electromagnetic theory to geophysical exploration. *Proceedings IEEE*, 67(7):979–1076, July 1979.

[Wait 1983] J. R. Wait. Resistivity response of a homogeneous earth with a contained vertical cylinder. *IEEE Transactions on Geoscience and Remote Sensing*, GE-21(1):109–113, January 1983.

[Wall 1980] D. J. N. Wall. Methods of overcoming numerical instabilities associated with the T-matrix method. In V. K. Varadan and V. V. Varadan, editors, *Acoustic, Electromagnetic and Elastic Wave Scattering — Focus on the T-matrix approach*, pages 269–286, Pergamon Press, New York, 1980.

[Wall 1987] D. J. N. Wall. A uniqueness theorem for the reduced wave equation governing the acoustic wave in a heterogeneous medium. *Quarterly Journal of Mechanics and Applied Mathematics*, 47:1–13, 1987. To appear.

[Wall *et al.* 1985] D. J. N. Wall, T. S. Yeo, and R. H. T. Bates. Inverse scattering and the null field method. In A. J. Devaney and R. H. T. Bates, editors, *Inverse Optics II*, pages 25–29, August 1985. Proceedings SPIE volume 558 (the Society of Photo-Optical Instrumentation Engineers).

[Wang 1979] K. Y. Wang, editor. Special issue on acoustic imaging. *Proceedings IEEE*, 67(4):452–675, April 1979.

[Ward and Saleh 1985] R. K. Ward and B. E. A. Saleh. Restoration of images distorted by systems of random impulse response. *Journal of the Optical Society of America A*, 2(8):1254–1259, August 1985.

[Waterman 1965] P. C. Waterman. Matrix formulation of electromagnetic scattering. *Proceedings IEEE*, 53(8):805–812, August 1965. Special Issue on Radar Reflectivity.

[Waterman 1969] P. C. Waterman. New formulation of acoustic scattering. *Journal of the Acoustical Society of America*, 45(6):1417–1429, June 1969.

[Wells 1977] P. N. T. Wells. *Biomedical Ultrasonics.* Academic Press, London, 1977.

[Wells 1987] P. N. T. Wells, editor. Special issue on medical imaging. *IEE Proceedings Part A*, 134(2):97–236, February 1987.

[Wells and Halliwell 1981] P. N. T. Wells and M. Halliwell. Speckle in ultrasonic imaging. *Ultrasonics*, 19(5):225–229, September 1981.

[Weston 1985] V. H. Weston. On the convergence of the Rytov approximation for the reduced wave equation. *Journal of Mathematical Physics*, 26(8):1979–1985, August 1985.

[Weston 1987] V. H. Weston. Factorization of the wave equation in higher dimensions. *Journal of Mathematical Physics*, 28(5):1061–1068, May 1987.

[Wilcox 1984] C. H. Wilcox. *Sound Propagation in Stratified Fluids.* Springer-Verlag, New York, 1984. Applied Mathematical Sciences volume 50.

[Williams 1984] J. E. C. Williams. Superconducting magnets for MRI. *IEEE Transactions on Nuclear Science*, NS-31(4):994–1005, August 1984. Special Issue on Physics and Engineering in Nuclear Magnetic Resonance Imaging.

[Wolf 1969] E. Wolf. Three-dimensional structure determination of semi-transparent objects from holographic data. *Optics Communications*, 1(4):153–156, September/October 1969.

[Wolf 1984] E. Wolf, editor. *Progress in Optics.* Volume XXI, North-Holland Physics, Amsterdam, 1984.

[Wolf and Nieto-Vesperinas 1985] E. Wolf and M. Nieto-Vesperinas. Analyticity of the angular spectrum amplitude of scattered fields and some of its consequences. *Journal of the Optical Society of America A*, 2(6):886–890, June 1985. Special Issue on Direct Problems in Propagation and Scattering.

[Wolf and Porter 1986] E. Wolf and R. P. Porter. On the physical contents of some integral equations for inverse scattering from inhomogeneous objects. *Radio Science*, 21(4):627–634, July-August 1986.

[Yagle 1986] A. E. Yagle. Differential and integral methods for multidimensional inverse scattering problems. *Journal of Mathematical Physics*, 27(10):2584–2591, October 1986.

[Yagle and Levy 1984] A. E. Yagle and B. C. Levy. Application of the Schur algorithm to the inverse problem for a layered acoustic medium. *Journal of the Acoustical Society of America*, 76(1):301–308, July 1984.

[Yagle and Levy 1986] A. E. Yagle and B. C. Levy. Layer-stripping solutions of multidimensional inverse scattering problems. *Journal of Mathematical Physics*, 27(6):1701–1710, June 1986.

[Zabreyko *et al.* 1975] P. P. Zabreyko *et al.* *Integral equations — a reference text.* Noordhoff, Leyden, 1975. Translated and edited by T. O. Shaposhnikova and R. S. Anderssen and S. G. Mikhlin.

[Zeeberg *et al.* 1985] B. R. Zeeberg, J. F. Soucaille, R. Carson, S. Bacharach, M. V. Green, and S. M. Larson. An efficient algorithm for reconstruction of SPECT images in the presence of spatially varying attenuation. *IEEE Transactions on Nuclear Science*, NS-32(2):1190–1197, April 1985.